Searching for sense in the library of Babel

Smith, Nick G. C.

Award date:
1999

Awarding institution:
University of Bath

Searching for Sense in the Library of Babel

submitted by Nick G.C. Smith

for the degree of PhD

of the University of Bath

1999

ProQuest LLC
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106-1346

Summary

The concept of genomic anatomy reduces the various relationships between nucleotide sequences within the genome to three fundamentals: the numbers of repeats, the levels of identity between repeats, and the relative positioning of repeats. In this thesis I consider whether genomic anatomy makes evolutionary sense, in other words is genomic anatomy the result of the deterministic processes of mutation and selection? To this end I have attempted to answer problems concerning the evolution of genomic anatomy at a number of scales. How can one explain variation in ploidy levels and life cycles, and hence the evolution of haplodiploidy? What role has polyploidisation played in vertebrate evolution? How do multiple copies of genes evolve? Is concerted evolution adaptive? Why are some genes arranged in homologous clusters? Given that most organisms are diploid why are some genes within a diploid cell expressed from only one copy, and how does such monoallelic expression affect the evolution of the genes involved? Why do genes evolve at different rates? Is selection strong enough to affect synonymous or intronic sites in mammals? The answers to these questions are rarely clear, and for the most part depend on the basic parameters of population genetics, about which we still know very little. The resolution of problems of genomic anatomy would appear to depend on an improved understanding of these fundamental genetic phenomena.

Acknowledgements

I would like to dedicate this work to three teachers who have inspired me: to my Grandad for early encouragement and the gifts of many books; to Dr. Gregory who persisted in teaching me biology at school even when I fell asleep at my desk from too much rowing; and to Laurence Hurst for his testable hypotheses, helpful supervision and generosity. I also take this opportunity to thank those with whom I have collaborated to produce the work in this thesis: Laurence Hurst, Clair Brunton, Rob Knight and James Randerson. Finally, I thank all my friends and family, especially the one whose purpose in life seems to be the disruption of my work.

Contents

Chapter 1. Searching for Sense in the Library of Babel

I should perhaps start with an explanation for the rather unusual form of this thesis. Rather than pursuing a single goal, I have worked on a number of projects over the last three years. These projects have yielded ten published papers and two manuscripts accepted for publication. But the subjects are rather diverse, and it would be disingenuous to suggest that all the work was conceived as part of a grand scheme. However, I need to present some sort of coherent story in this thesis, and so I shall propose a paradigm for explaining the links between the different subjects I have tackled. Therefore my thesis consists of largely independent research papers prefaced by introductions which should help to place them in context.

In this the first chapter I explain the paradigm I have adopted: what is genomic anatomy, and how might one make evolutionary sense of genomic anatomy? In Chapters 2 and 3 I consider genomic anatomy at the level of the haploid genome. How can one explain variation in ploidy levels and life cycles, and hence the evolution of haplodiploidy? What role has polyploidisation played in vertebrate evolution?

In Chapters 4 to 6 I turn my attention to genomic anatomy at the scale of the gene. How do multiple copies of genes evolve? Is concerted evolution adaptive? Why are some genes arranged in homologous clusters? Given that most organisms are diploid why are some genes within a diploid cell expressed from only one copy, and how does such monoallelic expression affect the evolution of the genes involved? Why do genes evolve at different rates?

In Chapters 7 and 8 I move down to the finest level of genomic anatomy, the level of the nucleotide within a gene. Is selection strong enough to affect synonymous or intronic sites in mammals?

In the Chapter 9 I review my findings in the context of the central aim of the thesis: can we make evolutionary sense of genomic anatomy? It seems that the solution to this problem lies in an improved understanding of the fundamental evolutionary processes.

Finally, I present those research papers which do not fit into the paradigm of genomic anatomy. How can sequence analysis methodology affect results in molecular evolution? Can one predict the evolutionary dynamics of selfish elements?

Genomes and libraries

My work is concerned with the evolution of genomes. But in order to study genomes one needs a descriptive framework, which I shall illustrate by comparing genomes and libraries. The similarities between genomes and libraries derive from the fact that both structures store information: they are both databases. By analogy with the familiar example of the library, I hope to show that the organisation of information within a database can be described in three terms; (1)

identity: the classification of units of information; (2) *copy number*: the number of copies of the same (full identity) or similar (partial identity) units of information within the database; and (3) *arrangement*: the organisation of the units information relative to each other. Both the *architecture of a library* and the *anatomy of a genome* can be defined in these terms.

So let us consider the human genome and Library of Babel described by Jorge Luis Borges (1941). (Dennet (1995) has also used Borges' story to explain evolution, although in a different way. I was unaware of Dennet's work when I wrote the first draft of this thesis.) Both the genome and the library contain enormous quantities of information. The Library of Babel "is *total* - perfect, complete, and whole - and its bookshelves contain all possible combinations of the twenty-two orthographic symbols." It contains about 10^{152} books, more than enough for each particle in our universe to have as many books each as there are particles in our universe. Although a genome does not contain all knowledge, with the human genome consisting of a mere 3 billion nucleotides, it does constitute the entire hereditary material of an organism, in the form of a precise nucleic acid sequence as well as additional epigenetic information. This is not to say that the genome contains all the information required to make an organism, as Lewontin (1993) and Dennet (1995) have stressed. Both a library and a genome need "readers" for the information to be used. In addition, and rather fortunately for evolutionary biologists, the genome carries information in the form of a record of its evolutionary history (Goldstein and Harvey 1999).

Identity in the Library of Babel can be defined at a number of levels: galleries, books, lines, even individual characters. Although the library is defined as possessing a single copy of every possible book, *copy number* variation can exist at lower units of organisation: "One book consisted of the letters M C V perversely repeated from the first line to last." The *arrangement* of the library is strictly hierarchical: "The universe (which others call the Library) is composed of an indefinite, perhaps infinite number of hexagonal galleries ... Each wall of each hexagon is furnished with five bookshelves; each bookshelf holds thirty-two books identical in format; each book contains four hundred and ten pages; each page, forty lines; each line, approximately eighty black letters."

Identity within the eukaryotic genome can be defined at a number of levels from the haploid genome to the individual nucleotide, with the corresponding *copy number* variation from polyploidy to microsatellites. The *arrangement* of the genome can similarly be described on a number of levels. At the cellular level the genome is divided into nuclear and cytoplasmic DNA. Nuclear DNA is organised into discrete linear structures called chromosomes, and consists of a number of copies of the haploid genome. Chromosomal organisation is exemplified by structural features such as centromeres and telomeres, and by chromosome banding patterns. At the regional level, isochores differ in nucleotide and transcriptional activity. Genes may be present in multiple copies, may be transcribed at different rates, and may be clustered with other genes. At the intragenic level, genes consist of introns and exons, and may themselves consist of repeated units. The notions of identity, copy number and arrangement can be used to describe the evolution of entire genomes (Huynen and Bork 1998).

But description is no substitute for explanation and so the next step is to ask why? In Borges' story the librarians concluded that the architecture of the library meant that there had to be an ultimate architect: "the universe, with its elegant appointments - its bookshelves, its enigmatic books, its indefatigable staircases for the traveller, and its water closets for the seated librarian - can only be the handiwork of a god." However, such explanations are unfalsifiable and hence unscientific (Ridley 1996). This thesis is an attempt to explain the anatomy of the genome in evolutionary terms, and that is why the title is "Searching for Sense in the Library of Babel". By evolutionary sense I mean deterministic evolutionary forces such as those of mutation or selection, as opposed to the stochastic evolutionary process of random drift (compare with the exclusively selectionist definition of Hurst 1999). This is not to say that random chance does not lead to discernible patterns of evolution (as in the neutral theory of Kimura 1983), rather that the products of random forces cannot be explained, they can only be described.

Consider, for example, patterns of base composition within a genome. Either selection or mutation can make sense of broad patterns of base composition such as isochores with high GC content. A selective explanation could be that high GC is favoured because it confers thermal stability to RNA structures. Alternatively, if mutation was biased towards G and C that would also make sense of the compositional bias. However, such high level explanations would not explain low level phenomena such as why a particular base is A rather than C, G, or T. At this level the base composition may be simply due to random chance, and hence the fact that the base happens to be A does not make evolutionary sense.

"Why?" questions in biology

Why is the genomic anatomy as it is? There are a number of ways of answering the question "why?", and Tinbergen (1963) has described the application of four such methods to ethology. These methods can be applied to the study of genomic anatomy. Two of Tinbergen's explanations, those of causation and development, can be combined in the context of genomic anatomy. These proximate answers describe how the genome is passed down cell lineages. The genome is the way it is because of genetic and epigenetic inheritance from the previous generation.

The selective explanation based on function, which suggests that a particular feature (character of genomic anatomy) is currently advantageous, is but a subset of the explanation from evolutionary history. The present generation possesses the feature as a necessary consequence of the feature being heritable and possessed by the previous generation. Thus the feature must have arisen at some point, then spread through the population, and then maintained until the present day. The primary importance of evolutionary explanations is illustrated by Dobzhansky's famous maxim: "Nothing in biology makes sense except in the light of evolution."

The approach used in this thesis is the approach from evolutionary history. Thus we can divide questions of existence into the two questions of origin and maintenance. Often these two questions have rather different answers, and there is no reason to suppose that simply because a

feature spread through the population for a certain reason then it is maintained for that reason. Indeed mutations may become fixed in the population as the result of genetic drift, in other words for no reason at all, and only later become selectively advantageous. Such a path is suggested by Wright's (1931) shifting balance theory.

Definitions of adaptation

Owing to a lack of information concerning mutational processes, as well as the generally held belief that most mutation processes are random, I have focused on the possible effects of selection on genomic anatomy. The question then becomes how much genomic anatomy is due to selection, and how much to random processes.

What is required of a selective explanation: how can we term a particular feature of genomic anatomy an adaptation? The term "adaptation" has a number of different meanings (see Futuyma 1998; Ridley 1996; Rose and Lauder 1996). The concept of adaptation used in this thesis is that of Futuyma (1998): "a feature is an adaptation for some function is it has become prevalent or is maintained in a population (or species, or clade) because of natural selection for that function". Thus a feature must have been subject to selection at some point in its evolutionary history, either positive selection to explain the initial evolution, or purifying selection to explain the maintenance of the feature. This is a very broad definition, but it is chosen in order to make adaptation consistent with evolutionary sense.

Futuyma's definition of adaptation is helpful because it encompasses a number of alternative definitions of adaptation and related concepts. Current selection is consistent with Reeve and Sherman's (1993) ahistorical definition: "An adaptation is a phenotypic variant that results in the highest fitness among a specified set of variants in given environment." Past selection allows an historical definition, as given by Harvey and Pagel (1991): "for a character to be regarded as an adaptation, it must be a derived character that evolved in response to a specific selective agent". Futuyama's definition also copes with the troublesome concepts of preadaptations, spandrels and exaptation (Gould 1997; Gould and Lewontin 1979). Thus a spandrel arose as the necessary consequence of selection and is thus an adaptation.

An adaptation is a product of natural selection. So for certain features of genomic anatomy to be adaptations, the fundamental requirements of natural selection must be fulfilled. Heritable variation must be generated through mutation, and selection must discriminate between variants. But how can we discern adaptations? A popular approach is the argument from design.

Design arguments

The Argument From Design (AFD) has a long history, all the way back to Aristotle (Amundson 1996; Barrow and Tipler 1996; Lauder 1996). The AFD assumes that complex

structures do not come about by chance, and that therefore some sort of deterministic process, either a creator or selection, is required.

The classical AFD supposes that the existence of a complex structure supposes an intelligent creator, a "final cause". Paley's famous mechanistic analogy is that the presence of a watch implies the existence of a watchmaker. Following the appearance of the *Origin of Species*, the Darwinian advocate Thomas Huxley announced the end of such teleological reasoning: "That which struck the present writer most forcibly on his first perusal of the Origin of Species was the conviction that teleology, as commonly understood, had received its death-blow at Mr. Darwin's hands." However, far from being rejected, the AFD was incorporated as an important part of the Darwinian paradigm. In modern terms then, a complex design is taken to imply a selective function and hence an adaptation, although there is some debate as to the meanings of the key terms of design, function and adaptation (Lauder 1996). As formulated by Williams (1992) the modern AFD is that "Adaptation is demonstrated by observed conformity to *a priori* design specifications."

The modern AFD can be attacked on the basis of philosophical considerations. Such problems stem from the idea that there are limitations to our reasoning about the real world. David Hume argued against the anthropocentric bias inherent in the AFD: "we are guilty of the grossest, and most narrow partiality, and make ourselves the model of the Universe ... What peculiar privilege has this little agitation of brain which we call thought, that we must make it the model of the whole universe." It is certainly true that unthinking anthropocentricity can lead to ridiculous adaptive explanations, such as this gem from Bernadin de Saint-Pierre (see Barrow and Tipler 1996): "Dogs are usually of two opposite colours, the one light and the other dark, in order that, wherever they may be in the house, they may be distinguished from the furniture, with the colour of which they might be confounded".

The application of the modern AFD raises a number of biological problems (Lauder 1996). The mechanistic analogy used by Paley may not apply to biological structures: there may be little correlation between structure and function. Even if structure and function are linked, we may not understand the causal relations, in other words we may not know which design criteria are important *a priori*. Even if there are *a priori* reasons for supposing that genomic anatomies differ in function, identifying the critical design features in a structure of such vast complexity (3 billion nucleotides, structured at a number of levels) will be difficult. Further problems are raised if one attempts to discern exactly what sub-feature selection acted on: complex features cannot be broken up into independent sub-features. Finally, it is difficult to use the AFD to provide an explanation from evolutionary history, rather than an explanation based on current function.

Hypothesis testing

If we reject the argument from design, how else can we study adaptation? The approach adopted in this thesis is that of classical hypothesis testing, as first proposed by Galileo and

Descartes. Descartes rejected the principle of final causation, the basis of the AFD, as anthropocentric and subjective, and doubted our ability to discern design: "the capacity of our mind is very mediocre, and not to presume too much on ourselves, as it seems we would do were we to persuade ourselves that it is only for our use that God has created all things, or even, indeed, if we pretended to be able to know by the force of our mind what are the ends for which he has created them."

In his *Discourse on Methods* Descartes (1637) introduced an approach to natural philosophy based on mathematical principles, and outlined the principles of hypothesis testing: "But I must also admit that the power of nature is so ample and so vast, and that these principles (*causes*) are so simple and so general, that I observe almost no individual effect without immediately knowing that it can be deduced in many different ways, and that my greatest difficulty is ordinarily to find in which of these ways the effect depends upon them; for to this end I know no other expedient but then to seek certain experiments which are such that their result not be the same if it is in one of these ways that the explanation lies as if it lies in another." Descartes' notion of "certain experiments" is equivalent to Galileo's idea of the critical experiment as the ordeal to which we must expose our hypotheses.

So in order to study adaptation we need to construct hypotheses, and then test them by performing the necessary experiments. Hypotheses can take a number of forms. They can be qualitative verbal arguments, or quantitatively explicit mathematical models. The predictions of mathematical models can be obtained by analysis or by simulation on computers. Analytical solutions may possess the security of mathematical proof, but simulations enable us to study problems which are mathematically intractable. Feynman (1966) has provided an eloquent defence of the simulation approach: "In the face of the lack of direct mathematical demonstration one must be careful and thorough to make sure of the point, and one should make a perpetual attempt to demonstrate as much of the truth as possible. Nevertheless, a very great deal more truth can become known than be proven". Hypotheses which are constructed in order to study adaptation need not invoke selection, indeed neutralist hypotheses are often more amenable to falsification than selectionist hypotheses. Kreitman (1996) has suggested that its ease of application ensures the importance of the neutral theory of molecular evolution, even if the theory itself is wrong.

There are three different ways of performing the "certain experiments" suggested by Descartes. The first method comes from the very process of formulating and modelling the hypothesis. Theoretical investigations act as thought experiments which determine whether a selective explanation could actually work. Of course, the validity of such thought experiments depends on the validity of the modelling they employ: is the model sufficiently close to reality, and are the parameter values appropriate? Often a lack of knowledge of realistic parameter values proves to be a stumbling block.

The second method is that of direct experimental manipulation followed by observation of the fitness consequences. Testing adaptation by experimental manipulation is useful for a number

of reasons (Sinervo and Basolo 1996). For example, current selection pressures can be assessed directly, and reconstructions of ancestral character states and ancestral environments allow the evolutionary history of the trait to be investigated. However, the direct experimental manipulation of genomic anatomy is difficult unless one can be sure that only the feature under investigation has been changed. There are also the associated problems of measuring small fitness effects and measuring fitness in a suitable environment.

The third method applies when we have no power to intervene in the process being studied. The correlational approach simply relies on multiple observations, in effect supposing that nature has performed a number of experimental manipulations. The theoretical basis of the correlational approach is given by the principle of concomitant variation, John Stuart Mill's fifth canon of experimental enquiry: "Whatever phenomenon varies in any manner whenever another phenomenon varies in some particular manner, is either a cause or an effect of that phenomenon, or is connected with it through some fact of causation." However, when one tests for correlations between evolutionary variables there is the "third variable problem", which can be eliminated in well designed artificial experiments but which cannot be eliminated from evolutionary data created by natural experiments (Brookfield 1997b). On this basis it has been argued that "the statistical significance of an observed correlation between traits is neither necessary nor sufficient for the inference of a causal connection between them" (Brookfield 1997b).

The correlational approach tests hypotheses which predict particular forms of concomitant variation between multiple observations. Tests of statistical significance assume the independence of multiple observations. But all living organisms are non-independent in the sense that they have evolved from a common ancestor. So the many species which use nucleic acid sequence as the hereditary material certainly do not constitute independent data. We can only conclude that one organism, the common ancestor of all extant species, used nucleic acids A single observation is certainly not significant. As Kundera (1984) points out: "*Einmal ist keinmal* ... What happens but once, says the German adage, might as well not have happened at all." We clearly need to account for the effects of phylogeny in order to justify the use of multiple data, and the comparative method (Harvey and Pagel 1991) has been developed to deal with this problem.

Comparative genomics aids our understanding of genomic anatomy in a number of ways (Clark 1999). The comparison of different genomes allows the identification of conserved linkage groups of genes and conserved functional regions within genes. The use of orthology relationships allows the extrapolation of results from model organisms, thus aiding the analysis of polygenic inheritance, the identification of novel genes and the elucidation of gene function.

The limits of selection

The notion of adaptive optimality is often ridiculed as Panglossian, after Voltaire's caricature of Leibniz's view of the harmony of nature: "All is for the best in the best of all possible worlds." A weaker formulation of the ubiquity of adaptation (in the sense used here) is Maynard

Smith's version of Dobzhansky's maxim: "Nothing in biology makes sense, except in the light of selection". However, there are several reasons to believe that selection is not all powerful, and even more reasons to suggest that optimality is likely to be rare. Such arguments do not constitute weaknesses of the theory of evolution by natural selection, but are instead strengths, a point made by the Darwinian advocate Asa Gray in 1876: "Darwinian teleology has the special advantage of accounting for the imperfections and failures as well as for the successes". I shall review the limitations to selection by dividing them into four classes: levels of selection, random drift, constraints, and spatial and temporal effects.

Levels of selection

As defined by Frank (1996), "An adaptive system is a population of entities that satisfy the three conditions of natural selection: the entities vary, they have continuity (heritability), and they differ in their success". If we consider the natural world, adaptive systems exist at a number of nested levels: genes, cells, individuals and groups of individuals. Selection will not necessarily act identically on all replicating units, which results in conflict. In Dawkins' (1990) terminology, conflicts arise because replicators differ in their desiderata lists. The conditions for conflict can deduced using Price's (1970) concept of fitness covariance. Negative fitness covariance between two replicators means that a change which leads to a gain in fitness for one replicator will diminish the fitness of the other replicator, and thus the two replicators are in conflict over fitness maximisation (Partridge and Hurst 1998).

When selection acts differently on different replicators it is impossible to say what state is optimal. Conflict can occur at all possible levels, though the traits under selection are always encoded by genes. The existence of conflict at a lower level engenders costs at a higher level, which means that selection (at the higher level) then favours genes which can reduce or prevent conflict.

Intragenomic conflicts can arise when alleles at a locus differ with respect to gene level fitness. An allele can spread if it has higher gene level fitness than its competitors, even if it reduces individual fitness. Such genetic conflict reduces individual fitness, and so a suppressor of conflict at another locus is then selectively favoured on the basis of individual level selection. Thus genetic conflict exists when "the spread of one gene creates the context for the spread of another gene, expressed in the same individual, and having the opposite effect" (Hurst *et al.* 1996a). A gene which spreads despite its cost to the individual is commonly termed a selfish gene, although all genes are selfish in the sense that they are selected to maximise their own fitness (Dawkins 1976), and so the term ultraselfish might be considered preferable.

Conflict can occur between any levels of selection. For example, Hamiltonian (1967) sex ratios arise from selection for Fisherian (1930) sex ratios between individuals within groups and selection for increased female bias between groups. The resultant Hamiltonian sex ratio is optimal in neither regard.

The concept of levels of selection is intimately related to that of the hierarchical organisation of biological entities. Maynard Smith and Szathmary (1995) have described the major transitions in evolution at which both biological complexity and the numbers of levels of selection increased. A major transition occurs when "entities that were capable of independent replication before the transition can replicate only as part of the larger whole after it". These major transitions are (1) the origin of chromosomes, (2) the origin of eukaryotes, (3) the origin of sex, (4) the origin of multicellular organisms, and (5) the origin of social groups. Of direct relevance to my thesis is the fact that the first three transitions are represented by changes in genomic anatomy, as well as increased conflict: (1) groups of genes in linkage, (2) nuclear and cytoplasmic inheritance, and (3) sex chromosomes and sex-specific inheritance. Maynard Smith and Szathmary (1995) have stressed the importance of central organisation and the division of labour as benefits to hierarchical organisation which may have driven the major transitions. The focus here is instead on the evolution subsequent to and contingent upon the major transitions.

The ubiquity of levels of selection seems to imply that conflict rather than cooperation should be universal. Why then does conflict appear to be the exception rather than the rule? A number of processes serve to reduce conflict, as in the following examples: suppression of conflict by "parliaments" of cotransmitted replicators, evolution of the genetic system to prevent conflict, and "tit for tat" strategies which reduce the benefits of selfish failures to reciprocate altruism (Hurst *et al.* 1992; Leigh 1991).

Random drift

Random genetic drift affects gene frequencies in finite populations, which means that changes under very weak selection are effectively neutral (Crow and Kimura 1970). Random genetic changes can arise from gamete sampling and from stochastic changes in selection intensity (Crow and Kimura 1970). Even mutations under relatively strong selection pressures stand only a slim chance of fixation, about twice the selective coefficient (Haldane 1927).

Another effect of finite populations is that the number of excess progeny is limited, which affects the number of loci upon which selection can act independently and simultaneously (Haldane 1957). The greater the number of excess progeny, the greater the amount of selective death that can occur.

Hill and Robertson (1966) found that selection at one locus reduces the probability of fixation of advantageous mutations at a second linked locus, an effect analogous to an increase in random drift. This Hill-Robertson effect is caused by all linked mutations affected by selection, both advantageous mutations via selective sweeps (Maynard Smith and Haigh 1974) and deleterious mutations via background selection (Charlesworth *et al.* 1993).

Constraints

The importance of constraints in evolution is a matter of considerable controversy. A number of different definitions are in use. In molecular evolution, a site is said to be constrained if stabilising selection acts upon it. In palaeontology, constrained changes are those which didn't occur. A strong evolutionary definition is that constraints are absolute blocks to the efficacy of natural selection: if a character is constrained then directional selection favouring that character is not possible (Laurence Hurst, pers. comm.). Strong constraints provide an absolute block to the attainment of optimality, by definition. A weak evolutionary definition is that a constraint is a partial block to natural selection: the effectiveness of directional selection is reduced by constraints. If environmental changes continuously move the optimum, then the reduced rate of approach to the optimum caused by weak constraints may prevent attainment of optimality.

Explanations based on strong constraint are controversial because they are intellectually lazy. It is easy to suppose that we do not observe a character because that character quite simply could not evolve. It is much harder to show that the character could have evolved but did not due because of subtle selective effects. A belief in widespread strong constraints is a belief based on ignorance.

Pleiotropy may act as a weak constraint. Pleiotropy occurs when a single mutation has multiple phenotypic effects which have differing fitness effects, some advantageous and some deleterious. The mutation will only spread if the good effects outweigh the bad ones. Thus seemingly potential increases in fitness are never realised. However, if the optimum phenotype can be obtained pleiotropy may improve the efficacy of stabilising selection (Wagner 1998b; Waxman and Peck 1998), although the loss of variation may itself prove a stumbling block to future adaptive change.

The genetic phenomena of penetrance, redundancy and epistasis may all be considered as weak constraints. Incomplete penetrance means that the full force of selection is not realised in diploids. Redundancy applies to both haploids and diploids and has the same effect as incomplete penetrance of reducing the power of selection. Epistasis can present a severe obstacle to the evolution of two alleles which are deleterious on their own but advantageous when together, although genetic drift in small populations may be able to overcome such obstacles (Haldane 1932).

The laws of physics and chemistry are clearly strong constraints. No mutations can enable organisms to escape such fundamental laws, and conversely we do not need to explain those characters which derive from such laws, a point missed by Dr. Pangloss: "Our legs are just long enough to reach the ground". Williams (1992) has illustrated this argument by showing that we do not need an adaptive explanation for why a flying fish returns to the water.

The "laws" of development and genetic systems may also act as strong constraints. Such developmental constraints may be considered as examples of contingent irreversibility (Maynard Smith and Szathmary 1995). Once a higher level of biological organisation is attained, then the

independent replicating ability of the lower level units will be lost through disuse and mutational decay. Thus genomic imprinting, a phenomenon dependent on the existence of separate sexes, acts in mammals as a barrier to the ancestral parthenogenetic mode of reproduction (Moore and Haig 1991).

Temporal and spatial effects

The appropriate selectively favourable mutations may be theoretically possible, but selection is powerless until such a mutation actually appears in the population. Even when selectively favourable mutations have occurred their spread to fixation will take many generations (Crow and Kimura 1970). The continuous generation of deleterious mutations and the time lag before loss means that deleterious mutations are present at a selection-mutation equilibrium. The lag to appearance and fixation of new mutations is important if selection pressures are constantly changing as the abiotic and biotic environments change.

Gene flow causes a spatial averaging of selection pressures (Kirkpatrick 1996). Selection pressures may vary over scales smaller than the habitat of a single organism, and thus migration prevents local adaptation.

How powerful is selection?

The arguments presented above suggest that selection has its limits and indicate those factors which might affect the efficacy of selection, but the quantitative importance of selection remains a matter of controversy. Differences of opinion with regard to the relative importance of selection versus random drift underpin both the controversy between Fisher (1930) and Wright (1932) and the more recent debate in molecular evolution between the neutralists (Kimura 1983) and the selectionists (Gillespie 1991).

I can think of three reasons why one might believe that selection is not the dominant cause of evolution. Firstly, one might reasonably invoke the strength of those processes which limit the efficacy of selection.

Secondly, one might object on the grounds of the supposed deterministic and fatalist implications of selection, as voiced by Kundera (1984): "Chance and chance alone has a message for us. Everything that occurs out of necessity, everything expected, repeated day in and day out, is mute". This is a philosophical rather than scientific point, but it may well lead some to believe in processes other than selection.

Thirdly, one might find it hard to believe that selection could possibly act on certain, seemingly trivial, characters. This is an argument based on our own ignorance, and is therefore to be mistrusted on both logical and historical grounds, as emphasised by Ridley (1996): "One character after another that had been written off as non-adaptive has turned out, following proper analysis, to be controlled by natural selection". In the context of the evolution of genomic

anatomy, translation selection on codon usage is now supported by a wealth of evidence, despite the fact that synonymous mutations were once considered prime candidates for selective neutrality.

Chapter 2. The evolution of life cycles

Ploidy level variation

The ploidy level of a cell is given by the number of copies of the haploid genome. It is possible to imagine a large number of alternative life cycles defined by different ploidy levels. Hence life cycle evolution can be viewed within the framework of genomic anatomy by considering whether the variation in ploidy copy number makes sense.

Sexual organisms are characterised by the "alternation of generations" because the ploidy level is halved at meiosis and doubled at syngamy (Mable and Otto 1998; Valero *et al.* 1992). There appear to be four main sexual life cycles characterised by an alteration between haploidy and diploidy (as defined by Mable and Otto 1998). (1) Diplonty: somatic development in diploid phase. (2) Haplonty: somatic development in haploid phase. (3) Haploid-diploidy: somatic development in both phases. (4) Haplo-diploidy: female somatic development in diploid phase, male somatic development in haploid phase. Although the near ubiquity of diplonty in metazoan taxa has lead to a belief in the superiority of diplonty and thus a focus on the advantages of diplonty (Mable and Otto 1998), there exists much variation in life cycles, both between higher order taxa (Lewis 1985) and within lower order taxa such the algae (Klinger 1993; Lewis 1985). There is no such necessity for an "alteration of generations" in asexuals, although ploidy cycles have been found (Kondrashov 1994).

I have studied the evolution of haplodiploidy, but in order to place the work in context I shall describe theoretical work on the relative merits of all four life cycles.

Haplonty versus diplonty

The evolution of life cycles and ploidy levels has been most thoroughly investigated with respect to haplonty versus diplonty. In order to illustrate the many potential differences between different life cycles, I shall summarise some of the major findings. There are three principle classes of arguments: genetic arguments concerning asexuals, genetic arguments concerning sexuals, and ecological arguments concerning both sexuals and asexuals.

Genetic arguments for asexuals

Ploidy levels affect the equilibrium genomic mutational load (Crow 1958) due to deleterious mutations. The mutation load due to germline mutations can be generalised to asexual N-ploids, in which case the load is given by $1 - e^{-NU}$ independent of epistasis, where U is the mutation rate per haploid genome (Kimura and Maruyama 1966; Kondrashov 1982). Thus

haplonty minimises such load. Haplonty is also favoured by intraorganismal selection between cells which contribute to the germline (Otto and Orive 1995). However, diplonty may be favoured if somatic mutations are partially recessive and sufficiently important (Orr 1995).

If one considers advantageous mutations then heterozygote advantage provides a clear advantage for diplonty over haplonty, since haploids cannot enjoy the benefits of overdominance (Crow and Kimura 1965) and asexuals do not suffer from segregational load (Crow and Kimura 1970). When populations are small so that favourable mutations arise rarely, diplonty allows a higher rate of novel adaptation than haplonty if advantageous mutations are partially dominant. But if populations are large so that favourable mutations appear fairly often, then haplonty is always favoured as long as advantageous mutations are not overdominant (Orr and Otto 1994).

Genetic arguments for sexuals

In outbred sexuals, the deleterious mutation load of diplonts is twice that of haplonts unless all mutations are completely recessive (Haldane 1937; Muller 1950). Inbreeding reduces the advantage of haplonty (Crow and Kimura 1970). When one considers genomic mutational load in sexuals then epistatic interactions become important (Kimura and Maruyama 1966). Synergistic epistasis can reduce the advantage of haplonty (Kondrashov and Crow 1991), and truncation-like selection can favour diplonty if dominance is less than ¼ and if the mutation rate per haploid genome is greater than one (Kondrashov and Crow 1991).

Contrary to the predictions of load arguments, a number of modifier analysis studies have shown that the transition from haplonty to diplonty can be favoured as a result of selection on genes outweighing selection on populations (Bengtsson 1992; Otto and Goldstein 1992; Perrot *et al.* 1991). Linkage, penetrance and the strength of selection all determine whether haplonty or diplonty is favoured. Haploid-diploidy is never favoured in such models. Also in contrast to load arguments, inbreeding favours haplonty rather than diplonty under modifier analysis (Otto and Marks 1996).

Intraorganismal selection can reduce the mutation load in sexuals and provide haplonts with the additional advantage of an effective reduction in mutation rate (Otto and Orive 1995). Somatic mutation can provide an advantage for diplonty, which may explain why developmentally complex organisms tend to be diplontic (Bell 1982; Orr 1995).

Could advantageous mutations provide an advantage to diplonty in sexual organisms? Heterozygote advantage may provide an advantage to diplonty, but seems unlikely to explain the initial evolution of diplonty (Goldstein 1992). With regard to the rate of novel adaptation, diplonty doubles the advantageous mutation rate but incomplete penetrance reduces the fixation rate of advantageous mutations in diplonts. High recombination rates reduce this cost of diplonty (Orr and Otto 1994).

Ecological arguments

Haploid cells tend to be smaller than diploid cells and thus enjoy a lower nutrient demand (Lewis 1985) as well as a higher rate of nutrient transport (Cavalier-Smith 1978; Mable and Otto 1998). Cavalier-Smith has extended such arguments to suggest that large complex organisms are K-selected for large cell size and hence diploidy, while simple unicellular organisms are r-selected for small cell size and hence haploidy (Cavalier-Smith 1978). Some unicellular organisms may be so large that multiple copies of the genome are required for cellular functions (Kondrashov 1994; Raikov 1982).

Alternative ecological arguments in favour of diplonty include the ability of diplonty to generate genetically diverse gametes (Bell 1997), a reduction in the mixing of vertically transmitted parasites (Hurst 1990), and more subtle control of development and gene regulation (Lewis 1985).

Why haploid-diploidy?

Many of the arguments given above favour haplonty under some parameter values and diplonty under other parameter values. But how can one explain the existence of haploid-diploid life cycles?

A variety of environments might favour two distinct phenotypes, with ploidy level providing a ready means to a heteromorphic life cycle (Mable and Otto 1998). Hurst and Nurse (1991) have suggested a number of arguments in favour of asexual ploidy cycles similar to those considered by Kondrashov (1994). High ploidy levels are favoured because large cells are better predators, less vulnerable prey, and more resistant to DNA damage. On the other hand, small haploid cells may respond better to nutrient stress, provide a greater number of survival units, and be able to grow faster. The logic of this hypothesis is dependent on the assumption of a constraint linking ploidy levels and cell size. Szathmary *et al.* (1990) have suggested that asexual ploidy cycles might be driven by temporal changes in the level of DNA damage.

Spores are reproductive propagules adapted for dispersal, while gametes are sexual propagules adapted for fusion. Such reproductive considerations, rather than any direct influence of ploidy level, appear to affect the diploid and haploid forms in brown seaweeds: microthalli are gametophytes while macrophalli are sporophytes (Bell 1997).

If viability rates are constant for both haploid and diploid sexual organisms, then either haplonty or diplonty is favoured, never haploid-diploidy. However, if viability decreases as a linear function of time at a constant ploidy level, then haploid-diploidy can be stably maintained. This result is not qualitatively affected by selection on viability loci; haploid-diploidy is possible both with overdominant selection and with disruptive selection (Jenkins 1993).

The evolution of haplodiploidy

Having dealt with haplonty, diplonty and haploid-diploidy, I now turn to haplodiploidy in which males develop from unfertilised eggs while females develop from fertilised eggs. There are also some life cycles in which males develop from fertilised eggs but, just as in haplodiploidy, transmit only the maternal genome to the next generation. Such life cycles have been termed paternal genome loss, parahaploidy, and male haploidy (Bull 1979), though male genetic haploidy may be preferable to the latter term (Goldstein 1994). In parahaploidy, the paternal genome may be excluded from all cells after fertilisation, or the paternal genome may persist in the somatic tissue but fail to be passed on to the sperm (Bull 1979). Different forms of parahaploidy may represent a series of ancestral stages in an evolutionary progression towards complete eradication of the paternal genome (Nur 1980).

Haplodiploidy appears to have arisen from diplonty several times, although the exact number is disputed (at least eight times (Oliver 1971), only seven or eight times (Hartl and Brown 1970), or at least eleven times (Borgia 1980)). Haplodiploidy is found in the monogonant rotifers, bees and wasps, thrips, sawflies, ticks and mites, bark beetles and scale insects. Parahaploidy is found in diaspidid scale insects, lecanoid scales and sciarid flies.

In comparison to thelytoky, the other major alternative to diplonty, transitions to haplodiploidy appear to have been rarer. However, haplodiploidy appears to be a more stable, with the result that haplodiploid groups comprise large taxonomic groups whereas thelytokous groups are often of low taxonomic rank (Borgia 1980).

Haplodiploidy versus diplonty

What are the differences between the haplodiploid and diplontic life cycles? Two important issues are those of sex specific copy number and parent of origin effects. In these respects the evolution of haplodiploidy relates to two topics I consider later in this thesis: X-linked genes and genomically imprinted genes.

In terms of copy number and genetic transmission the population genetics of haplodiploids and X-linked genes in diplonts are fundamentally similar (Hedrick and Parker 1997). One difference lies in the level of genomic anatomy at which the copy number variation is expressed: male haplodiploids have only one copy of the genome, whereas male diplonts have only one copy of the X chromosome (assuming males to be the heterogametic sex).

Parent of origin effects apply both to haplodiploidy and genomic imprinting in diplonts. In genomic imprinting the expression of an allele depends on the parent from which it was inherited, and both paternal and maternal silencing is found (Efstratiadis 1994). In haplodiploidy it is the transmission and expression of genes which is dependent on the parent of origin, with the paternal genome neither transmitted nor expressed.

The population genetics of haplodiploids (also X-linked genes) differ in a number of ways from the population genetics of diplonts. Selection is more effective in haploids, and this is reflected in the differences between haplodiploids and diplonts. Assuming similar allele frequencies in the two sexes and no dosage effects, the response to selection is about one third faster in haplodiploids (Hartl 1972). Advantageous mutations spread faster, and deleterious mutations are eliminated more effectively. Diplonty is however favoured in the case of overdominance, since heterozygote advantage cannot be enjoyed by haploids.

If one allows sex specific allele frequencies the average mutational load under outbreeding in haplodiploids is only three quarters that of diplonts (Werren 1993). The average load of haplodiploids obeys the Haldane-Muller rule of insensitivity to selection pressures, but the separate male and female loads do vary as both selection pressures and dominance relations change (Werren 1993). When penetrance is low and selection is strong, male load is high relative to female load because selection is only effective in haploids. But when penetrance is high, then selection acts equally in haploids and diploids, and female load is higher due to mutations appearing at twice the rate in males. These arguments assume no adaptive variation in the mutation rate, an assumption which does not appear to hold for the analogous case of X-linked genes, as discussed later.

Assuming equal numbers of males and females, the effective population size of a haplodiploid population is only three quarters of that of a diplont population of the same number of individuals. Since there is no recombination in males, the recombination rate of haplodiploids is two thirds that of diplonts. Both these factors combine to make random drift more important in haplodiploid populations. With regard to the evolution of sex and recombination, haplodiploidy might be thought to suffer from a loss of recombination, both in terms of the spread of advantageous mutations and the removal of deleterious mutations. Such arguments against haplodiploidy are weakened by the absence of recombination in male *Drosophila*, by the possibility of selection for higher female recombination rates, and by the idea that only a little bit of recombination is required for sex to be beneficial (Green and Noakes 1995).

Inbreeding and population structure affect haplodiploids and diplonts in many different ways: mutational load, the effectiveness of selection, sex ratio adjustment, sperm shortage, and sex determination are all involved. Many of these issues are intertwined, and also affect theories on the origins of haplodiploidy. Ignoring origins and comparative methods for the moment, it does appear that haplodiploids tend to be highly inbred and have female biased sex ratios (Hamilton 1967).

Inbreeding in haplodiploids differs from inbreeding in diplonts because of relatedness asymmetries and because haploid males cannot be inbred (Hedrick and Parker 1997). Studies of the mutation load shows that haplodiploids can deal with inbreeding better than diplonts (Werren 1993). Although diplonts actually do rather better from chronic inbreeding than haplodiploids in terms of reducing mutation load, haplodiploidy enables an easier transition from outbreeding to inbreeding. Under acute inbreeding, a sudden transition from outbreeding to obligatory sib-sib

mating, diploids suffer a severe increase in load, while female haplodiploids suffer a less extreme increase, and male haplodiploids actually enjoy a diminution of their load.

A number of potential problems exist with this analysis. The mating structure will affect the nature of selection on the sexes. Under sib-mating, weakened males will be protected from competition from stronger males (Borgia 1980), but this advantage to the males might adversely affect their sisters if weak males produce inviable sperm (Bull 1979). Haplodiploidy provides an escape from low population densities via inbreeding: a female can produce a male and then mate with him (Borgia 1980; Hartl and Brown 1970). One cost of inbreeding which will affect haplodiploids more than diplonts is the cost of sex determination (Werren 1993). If males are hemizygous and females are heterozygous at the sex locus, then inbreeding will cause the production of inviable diploid males.

Inbreeding promotes the evolution of female biased sex ratios (Hamilton 1967), although it is neither necessary nor sufficient (Charnov 1982). Under the levels of selection formulation of the Local Mate Competition model of sex ratio selection, it is group selection which favours female bias. Selection on groups increases as the variance between groups increases, and so small groups with their greater sampling errors cause greater female bias. Inbreeding will tend to be more common in small groups, and so inbreeding covaries with selection for biased sex ratios.

Whatever the reason for which female bias is favoured, the adjustment of sex ratio is thought to be easier in haplodiploids, where the female can choose sex on the basis of fertilisation, than in diplonts, whose genetic systems have evolved to prevent meiotic drive (Borgia 1980; Charnov 1982; Hurst 1992a). The favouring of female biased sex ratios also provides an advantage to the sex specific mutation loads of haplodiploids, since females reap the benefit of higher load in males (Werren 1993). It should be noted that the Fisherian argument does work for both diplonts and haplodiploids under panmixia, and so female biased sex ratios are not favoured in haplodiploids simply because females are diploid and males are haploid (Hartl and Brown 1970). However, Fisherian sex ratios depend both on the mating structure and the controlling element (Hamilton 1995). When control is zygotic, as opposed to maternal, then the intuitively obvious two to one female bias is favoured (Bull 1983). The potential importance of conflict over the sex ratio is shown by Haig's (1993) theory of the origin of parahaploidy which is discussed below.

Both Borgia (1980) and Brown (1964) have suggested that haplodiploidy may have evolved in order to prevent conflict rather than as a result of conflict. If males were initially heterogametic under diplonty, then the evolution of haplodiploidy would prevent meiotic drive. The force of this argument is weakened in the presence of inbreeding, since inbreeding itself reduces the potential for conflict by increasing fitness covariance between replicators.

The mutation load results given above were calculated on the basis of no dosage effects. In other words, males suffered no specific cost as the result of being haploid rather than diploid. A belief that haploidy will incur some disadvantage is prevalent in the literature (Borgia 1980; Hartl and Brown 1970), but if haplodiploidy can overcome this initial hurdle, then sex specific dosage

compensation would be selectively favoured (Hartl and Brown 1970). One immediate effect of male haploidy which might be advantageous is the consequent reduction in cell and organism size: Cavalier-Smith (1978) has suggested that haplodiploidy may be the result of selection for males to be smaller than females, which is consistent with the sexual dimorphism often found in haplodiploids. This argument appears weak when one considers giant fighting haplodiploid males (Hamilton 1979), and furthermore offers no explanation for those forms of parahaploidy where the paternal genome is present but silenced through heterochromatisation.

Haplodiploidy may have facilitated the evolution of eusociality (Crozier and Pamilo 1996), with all its attendant advantages of central control and division of labour (Maynard Smith and Szathmary 1995). Such an effect may well explain why haplodiploid species have done so well, but does not provide a satisfactory answer to the question of how haplodiploidy first evolved, assuming that the evolution of eusociality was subsequent to the evolution of haplodiploidy.

The origins of haplodiploidy and parahaploidy

The question of how and why haplodiploidy first evolved has been answered in a number of ways. The ancestral state is assumed to be diplonty, although this has only been confirmed in some cases (Goldstein 1994). Leaving aside Cavalier-Smith's (1978) cell size argument, and Whiting's (1945) race hybrid model, there seem to be three main classes of explanation for the transition from diplonty to haplodiploidy.

The first explanation is based on deleterious mutations. Using modifier analysis, Goldstein (1994) has shown that the diplonty to haplodiploid transition for outbred sexuals is permitted by a broader range of parameter values than is the diplonty to haplonty transition. The invasion of haplodiploidy is more permissive than the invasion of haplonty because recombination does not occur in haploid males. Recombination tends to favour diplonty in this system , and so haplodiploidy can sometimes invade even when penetrance is too low for haplonty to invade (low penetrance favours diplonty).

The second explanation makes use of the relatedness asymmetries of haplodiploidy. All of the genes of a haplodiploid male come from his mother, whereas only half of the genes of a diplontic male are maternally derived. If the control of the genetic system rests with the mother rather than the father, then haplodiploidy can spread because it favours the transmission of maternal genes. Early models of this "maternal transmission" argument assumed that haplodiploidy gained through the ability of females to produce males from unfertilised eggs (Brown 1964; Hartl and Brown 1970). Hartl and Brown (1970) found that for both environmental and genic control haplodiploidy fully displaced diplonty if haploid male fitness was high enough relative to the proportion of unfertilised eggs. Bull (1979) generalised the model by noting that the proportion of unfertilised eggs might well be under maternal control, and found that haplodiploidy evolved as long as haploid male fitness was more than half diploid male fitness. Therefore the maternal transmission argument rests on a two fold increase in the representation of maternal genes

in the gametes of haploid sons (Bull 1979). The maternal transmission argument is based on conflict, since the evolution of haplodiploidy is "driven" despite the negative fitness consequences at the level of the individual male.

Haig (1993) has suggested an alternative conflict based scheme which may explain the evolution of the different life cycles of the coccids (see Nur 1980). Males are XO and females are XX. The argument starts by noting that in the absence of panmixia both autosomal and X-linked genes "favour" female biased sex ratios although there is conflict between the X and the autosomes with regard to the extent of bias. The autosomes can respond to a driving X chromosome by suppressing drive, or they may join forces with the driver by evolving X linkage. Conflict over the sex ratio would then occur between individual females and the X-autosome driving maternal genome. The females might resolve this conflict by converting XX females to sons by inactivating the paternal genome. Different forms of inactivation provide explanations for the different forms of coccid parahaploidy and possibly even true haplodiploidy.

Haig's "historical narrative of shifting alliances, innovations and counter-strategies" appears plausible, and provides an attractive conflict based view of life cycle evolution. Hamilton (1993) has suggested that an iterative process of chromosome loss may provide a different conflict based explanation for the evolution of haplodiploidy. So why is the paternal chromosome set retained in parahaploidy? Although heterochromatic, the paternal chromosomes do appear to perform a specialised role which is necessary for both male viability and male fertility (Brown and Nur 1964). Alternatively, sex determination may rely on the paternal set of chromosomes.

Weak constraints (as defined in Chapter 1) are thought to hinder the initial evolution of haplodiploidy (Borgia 1980; Bull 1979; Hartl and Brown 1970; Whiting 1945). A number of preconditions must be overcome if true haplodiploidy is to evolve. The developmental block to unfertilised eggs must be overcome. Haploid eggs must develop into males rather than females. The haploid male must overcome the viability problems of gene dosage and the appearance of deleterious mutations previously masked in diploids. Finally, haploid males must be fertile, so spermatogenesis needs to work in haploids. Of all these weak constraints, the most severe appears to be the requirement that haploids are male, which places a restriction on the mechanism of sex determination (Bull 1979). The conflict based explanation of parahaploidy requires XO males and XX females, while both the deleterious mutation and maternal transmission explanations of haplodiploidy appear to require XX males and XO females (Borgia 1980).

The effect of inbreeding on the evolution of haplodiploidy

As described above, haplodiploids complete the transition to inbreeding more easily than diplonts, and haplodiploids may also gain under inbreeding through an improved ability to adjust sex ratios. But how does inbreeding affect the deleterious mutation and maternal transmission explanations of haplodiploidy?

Borgia (1980) has argued strongly in favour of inbreeding providing an advantage to the transition to haplodiploidy. However, it should be noted that the correspondence between inbreeding, sex ratio bias and haplodiploidy does not necessarily imply that inbreeding caused haplodiploidy. Hamilton makes this point in one of his seminal papers (1967): "Arrhenotoky is a mode of reproduction that readily permits the production of biased sex ratios, with control of the sex ratio normally dependent on the phenotype of the mother. Therefore it seems either that male-haploid organisms have found themselves pre-adapted for life in niches of the sort characterised by the (Local Mate Competition) model or that the evolution of male haploidy has actually accompanied, in several independent lines, an evolutionary trend to occupy such niches".

Inbreeding leads to a reduction in the heterozygote and gamete variability advantages of diplonty over haplodiploidy (Borgia 1980). The reduction of mutation load due to inbreeding might be expected to favour haplodiploidy under the deleterious mutation hypothesis (Goldstein 1994), in the same way that the transition from diplonty to haplonty is favoured by inbred mating systems (Otto and Marks 1996). The effect of inbreeding on the maternal transmission hypothesis is less clear. Borgia (1980) has claimed that inbreeding favours the evolution of haplodiploidy because weak males are sheltered from competition, but Bull (1979) has pointed out that female relatives of the weak males may suffer a fecundity cost of inbreeding, and also that maternal transmission is only favoured in heterozygotes which are rarer under inbreeding.

I have attempted to understand whether inbreeding does favour haplodiploidy by adapting Goldstein's (1994) model of deleterious mutations to incorporate both maternal transmission and inbreeding in the form of sib matings (see Research Paper 1). Simulations reveal that inbreeding favours the evolution of haplodiploidy under the deleterious mutation hypothesis. Haplodiploidy can spread within inbred populations under lower levels of penetrance than within outbred populations. However, inbreeding restricts the conditions under which haplodiploidy evolves as the result of a maternal transmission advantage. Under outbreeding haploid fitness needs to be more than half diploid fitness for haplodiploidy to spread, but under complete sib mating the requirement is that haploids are fitter than diploids.

These simulation results confirm intuitive arguments, but in addition permit an evaluation of the extent to which inbreeding affects the likelihood of haplodiploidy evolving. It is not known which of the two effects, deleterious mutations or maternal transmission, is stronger, but the effects of inbreeding on both models can be compared. It appears that the effect of inbreeding is greater under the maternal transmission model than under the deleterious mutation model, and thus inbreeding probably makes the evolution of haplodiploidy less likely.

Why male haplodiploidy?

One final question is why haplodiploidy always takes the form of male haploidy. Under LMC group productivity is limited by females, and so the fact that males bear the brunt of the

mutations favours male haploidy (Werren 1993). An associated point is that if female biased sex ratios are favoured but cannot be attained by diplonts, the transition to male haploidy is favoured by selection on the sex ratio.

Bull (1979; 1983b) has described the four theoretically possible systems of haplodiploidy. Two of these require paternal control, which seems unlikely, and of the remaining two forms, only male haploidy is sexual and hence likely to persist over evolutionary time (Bull 1979; Bull 1983; Harvey and Partridge 1984).

Research Paper 1. The evolution of haplodiploidy under inbreeding

Nick Smith (1999)

Heredity, in press.

Summary

Although haplodiploid organisms tend to be inbred, previous models of the evolution of haplodiploidy have assumed outbred populations. Here a model for the evolution of haplodiploidy is developed which incorporates sib mating, deleterious mutations generated by mutation, and fitness differences between haploids and diploids. Simulations of the model allow an assessment of the effect of inbreeding on the deleterious mutation and maternal transmission theories for the evolution of haplodiploidy. As expected from intuitive arguments, inbreeding favours haplodiploidy under the deleterious mutation hypothesis but disfavours haplodiploidy under the maternal transmission hypothesis. It appears that the effect of inbreeding is greater on the maternal transmission theory, and thus inbreeding may restrict the evolution of haplodiploidy.

Introduction

Sexual organisms are characterised by the "alternation of generations" because the ploidy level is halved at meiosis and doubled at syngamy (Valero *et al.*, 1992; Mable & Otto, 1998). Alternative sexual life cycles can be characterised by the timing of alternations between haploidy and diploidy. In diplonty somatic development takes place in the diploid phase, while in haplonty somatic development occurs in the haploid phase. In haplodiploid organisms the sexes differ in the ploidy level at which somatic development takes place: females develop as diploids while males develop as haploids. Diplonty is by far the most common life cycle found in animals, although there are two other common life cycles. One is asexual thelytoky, the evolution of which is addressed by the vast literature on the evolution of sex (see Hurst & Peck, 1996). The other is haplodiploidy, which appears to be a more stable and successful alternative to diplonty than asexuality (Borgia, 1980).

Although the evolutionary consequences of haplodiploidy have attracted a good deal of interest, especially in the context of social evolution (e.g. see Bourke & Franks, 1995), the evolutionary origins of haplodiploidy are still unclear (Whiting, 1945; Brown, 1964; Hartl & Brown, 1970; Bull, 1979; Borgia, 1980; Haig, 1993; Goldstein, 1994). The intention of this study is to discriminate between two hypotheses for the evolution of haplodiploidy from diplonty with respect to the effect of inbreeding. Although the number of independent origins of haplodiploidy is uncertain, it is generally agreed that haplodiploidy arose from diplonty (Hartl & Brown, 1970; Oliver, 1971; Borgia, 1980).

The maternal transmission hypothesis derives from the relatedness asymmetries of haplodiploidy. All the genes of a haplodiploid male come from his mother whereas only half of the genes of a diplontic male come from his mother. Thus a haplodiploidy modifier can spread in a diplontic population because of its enhanced transmission through the maternal germline. Haplodiploidy can spread even if haploid males are less fit than diploid males, and thus the maternal transmission hypothesis invokes a selfish gene which spreads despite reducing population fitness. Haplodiploidy is predicted to evolve in outbred populations as long as haploid male fitness is more than half diploid male fitness (Brown, 1964; Hartl & Brown, 1970; Bull, 1979). In this

study I shall a version of the maternal transmission hypothesis in which the modifier affects the transmission of the entire haploid genome, but it is possible that a similar process might act on a single chromosome. Indeed Hamilton (1993) has suggested that haplodiploidy might have evolved one chromosome at a time.

The deleterious mutation hypothesis suggests that a haplodiploidy modifier can spread in a diplontic population because of its indirect fitness consequences via selection on deleterious mutations (Goldstein, 1994). As with similar models for the transition between haplonty and diplonty (Otto & Goldstein, 1992), the spread of haplodiploidy is dependent on the strength of selection against deleterious mutations, the penetrance of deleterious mutations in diploids, and the level of recombination (Goldstein, 1994). Load arguments show that the mutational load in a haplodiploid population is lower than in a diplontic population (in outbred populations with no sex ratio bias the haplodiploid load is ¾ of the diplontic load). But the situation in an evolving population is more complex, as shown by modifier analysis. Purifying selection against deleterious mutations at viability loci is more efficient in haploids than in diploids if penetrance is incomplete, and so selection generates a linkage disequilibrium in which the haplodiploidy modifier associates with higher fitness viability allele (and similarly the diplonty modifier associates with lower fitness viability allele). In the absence of recombination, haplodiploidy is always favoured over diplonty. But if there is recombination then the linkage disequilibrium is reduced, which means that the cost of diplonty (association with lower fitness viability allele) is reduced. If the rate of recombination is high enough, then diplonty can actually be favoured over haplodiploidy because of incomplete penetrance since a diploid is less affected by a single deleterious mutation than a haploid.

Inbreeding appears to be associated with haplodiploidy (Hamilton, 1967). Even if it is assumed that this association is the result of a causal relationship rather than the confounding effect of a third variable, and that the perceived association is robust to the demands of the comparative method, it is still not clear in which direction the causality lies. Haplodiploids may well complete the transition to inbreeding more easily than diplonts (Werren, 1993). Borgia (1980) has stressed the alternative view that inbreeding eases the transition from diplonty to haplodiploidy.

Intuition suggests that the reduction of mutation load relative to panmixia under inbreeding might be expected to favour haplodiploidy under the deleterious mutation hypothesis in two ways.

First, inbreeding reduces the difference between diploid mutation load and haplodiploid mutation load: "diploid species appear to "benefit" more than haplodiploids from chronic inbreeding, in terms of a reduction in genetic load" (Werren, 1993). This effect reduces the cost of the transition from diplonty to haplodiploidy, since the cost derives from the exposure of the standing body of deleterious recessives present in diplonts. The second reason is based on modifier analysis rather than load arguments. Inbreeding reduces the effectiveness of recombination. As argued above, a reduction in recombination ought to favour haplodiploidy over diplonty, since recombination acts to favour diplonty by breaking up the association between the haplodiploidy modifier and the higher fitness viability allele. The effect of inbreeding has already been investigated for the transition from diplonty to haplonty under a variety of mating schemes, and as predicted inbreeding favours haplonty over diplonty (Otto & Marks, 1996).

However it also seems likely that inbreeding acts against the maternal transmission hypothesis for the evolution of haplodiploidy. Inbreeding reduces the proportion of heterozygotes, and the spread of the haplodiploidy modifier is favoured by transmission through females heterozygous at the ploidy modifier locus (Bull, 1979). The haplodiploidy modifier acts a selfish gene and can only spread by virtue of its improved transmission relative to the diplonty modifier. In the extreme case, if inbreeding is complete, then the two ploidy modifier alleles can never meet, and so the haplodiploidy modifier can never spread.

The purpose of this study is to address two questions relating to the evolution of haplodiploidy under inbreeding. First, are the intuitive arguments correct in predicting the effect of inbreeding on the maternal transmission and deleterious mutation hypotheses? Secondly, can we predict the overall effect of inbreeding on the evolution of haplodiploidy?

The model

Goldstein's (1994) model provides a scheme for incorporating both the maternal transmission and deleterious mutation hypotheses under panmixia. Using modifier analysis Goldstein was able to derive analytically the conditions under which a haplodiploidy modifier spreads in a diplont population at mutation-selection equilibrium. I have extended the model to include both random mating and sib mating. This change entails a considerable increase in the

complexity of the algebra. Under panmixia, the recursions operate on a total of eight gamete frequencies, but under sib mating recursions have to be performed on the 90 possible adult mating pairs. I have used a computer program to perform simulations by calculating the recursions automatically.

Consider two autosomal loci, one viability locus and one ploidy modifier locus. The two loci recombine at the rate r per meiosis. There are two alleles A and a at the viability locus. Allele A is wild-type with relative viability 1 in both homozygotes and hemizygotes (AA and A). Allele a is deleterious with relative viability $1-s$ in homozygotes and hemizygotes (aa and a) and $1-hs$ in heterozygotes (Aa). Thus the coefficient of selection against deleterious mutations is represented by s, and the penetrance of deleterious mutations in diploids is given by h. Mutation is unidirectional, from A to a at a rate μ per gamete.

At the ploidy modifier locus there are two alleles P and p. As there are two alleles at both loci there is a total of four gamete types: *Ap*, *ap*, *AP*, and *aP*. There are eight gamete frequencies which need to be specified in the model because gamete frequencies can differ between the sexes under haplodiploidy. Diploids produce gametes by meiosis, and haploid males are considered to produce sperm by a process analogous to mitosis. Diploid females are produced by the union of male and female gametes. The ploidy level of males depends on the genotype of their mother's egg at the ploidy locus. If the female gamete contains P then the unfertilized egg develops into a haploid male, but if the female gamete contains p then the egg is fertilized and develops into a diploid male. In other words, P makes males out of unfertilized eggs. The control of sex determination is assumed to maintain a 1:1 sex ratio. The relative viability of haploid males versus diploid males is given by haploid fitness f. The scheme under panmixia is shown in Fig. 1.

The breeding system is quantified by the variable *sib*, which represents the proportion of exclusively sib matings as opposed to random matings. The investigation of sib mating requires that recursions be performed on mating-pair frequencies rather than gamete frequencies. There are nine different male genotypes (*ApAp, Apap, ApAP, ApaP, apap, apAP, apaP, AP, aP*) and 10 different female genotypes (*ApAp, Apap, ApAP, ApaP, apap, apAP, apaP, APAP, APaP, aPaP*), so there is a total of 90 mating-pairs. For all mating-pairs the contributions of sib matings and random matings are combined in the proportions *sib* and 1 - *sib* respectively.

Under the random mating scheme the proportions of the offspring types of each mating-pair are pooled across all mating-pairs weighted by mating-pair frequencies. It is assumed that each brood produces an effectively infinite number of offspring so that stochastic variation can be ignored. Gamete production and recombination, gamete mutation and random gamete fusion are invoked before the pooling of offspring types. Adults die after they have produced offspring, so do not contribute themselves to the next generation. Then viability selection followed by normalization acts on the pooled offspring. Finally, random mating pairs are calculated by the products of the adult frequencies. This procedure carries the assumptions that mating is at random, that there is an effectively infinite number of adults, and that all adults mate.

Under the sib mating scheme new mating-pairs are generated within each mating-pair. Again an effectively infinite number of offspring within each brood ensures that all females are mated and that there is no stochastic variation. All the following processes therefore take place within the brood of each mating-pair: gamete production and recombination, gamete mutation, random gamete fusion to produce brood progeny, death of parents following reproduction, viability selection without normalization on offspring, and mating-pairs generated by the random union of surviving offspring within the brood. Following pooling across all mating-pairs the new mating-pair frequencies generated by sib mating are normalized.

Invasion by haplodiploidy

For each simulation, values were assigned to the following parameters: μ, r, h, s, f, and sib. The simulation began with the sole mating pair as *ApAp* - *ApAp* and was allowed to proceed until the diplontic population reached mutation-selection equilibrium. Thus in the case of $sib > 0$ inbreeding was chronic, rather than acute (see Discussion). Then the *P* allele was introduced in the form of a small proportion (0.001) of the *APAP* - *APAP* mating pair. The frequency of the *P* allele at the ploidy modifier locus was then calculated after 1000 and 2000 further generations, with the *P* allele frequencies at such times given by *P1* and *P2* respectively. If the *P* allele rose in frequency or had reached fixation ($\delta P = P2 - P1 > 0$ or $P2 = P1 = 1$) then the haplodiploidy modifier had invaded the population. This simulation approach is similar to the methods of modifier analysis.

Panmixia

Under panmixia ($sib = 0$) and no selection against deleterious mutations ($s = 0$ or $\mu = 0$) the critical value of f is ½, as found in previous models of the maternal transmission hypothesis (Bull, 1979; Goldstein, 1994). Above this value haplodiploidy will always invade and below this value haplodiploidy can never invade.

If f is set to ½ then it is possible to examine the critical values for the invasion of haplodiploidy under the deleterious mutation hypothesis. The recombination rate r was set to ½ in all simulations under the assumption that a single ploidy modifier will be freely recombining with most loci in a genome with several chromosomes. The mutation rate μ was set to 0.0001 in all simulations. The critical values of h and s were in close agreement with Goldstein's (1994) analytical results which were obtained by ignoring terms of the order of μ^2 (see Fig. 2). High penetrance favours haplodiploidy because then the advantage to diplonty of the masking of deleterious mutations in diploids is reduced. Low recombination also favours haplodiploidy (results not shown but see Goldstein, 1994) because mutation load effects which favour haplodiploidy over diploidy then increase in magnitude. The reduced recombination in haplodiploidy, due to the lack of recombination in haploid males, explains why the conditions for the invasion of haplodiploidy are marginally less restrictive than the conditions for the invasion of full haplonty (Goldstein, 1994). More powerful selection against deleterious mutations also favours haplodiploidy (see Fig. 2) since stronger selection increases the association between the haplodiploidy modifier and the higher fitness viability allele.

Inbreeding

When sib mating is introduced into the model ($1 \geq sib > 0$), interpretation of results becomes slightly more complicated because inbreeding affects both the maternal transmission and deleterious mutation arguments.

The effect of inbreeding on the maternal transmission hypothesis can be easily seen by removing selection against deleterious mutations ($s = 0$ or $\mu = 0$). As expected the critical value of

f rises as the proportion of sib matings increases (see Fig. 3). When all matings are between sibs ($sib = 1$) then haploid fitness must exceed diploid fitness ($f > 1$) for haplodiploidy to invade.

It is not possible to consider the deleterious mutation hypothesis in isolation, but it is possible to account for the effects of maternal transmission. For a certain level of inbreeding, the critical value of f gives a minimal rate of haplodiploidy invasion *inv* ($inv = \delta P$). Although *inv* should clearly be zero at the critical point, unlimited precision cannot be achieved in simulations. However, the critical value of f was estimated by trial and error to a precision of 0.001, and so the resulting values of *inv* were very small. When selection against deleterious mutations is invoked, f is set to its critical value for the current value of *sib* and the requirement for haplodiploidy invasion then becomes $\delta P > inv$ rather than $\delta P > 0$. Thus it is possible to measure the effect of inbreeding on the deleterious mutation hypothesis. As expected inbreeding facilitates the invasion of haplodiploidy according to the deleterious mutation hypothesis. For a given selective coefficient inbreeding reduces the minimum level of penetrance required for haplodiploidy to invade, once the effect of the maternal transmission hypothesis has been taken into account by adjusting f (see Table 1).

Discussion

The purpose of this study is to address two questions relating to the evolution of haplodiploidy under inbreeding. First, are intuitive arguments correct in predicting the effect of inbreeding on the maternal transmission and deleterious mutation hypotheses? Secondly, can we predict the overall effect of inbreeding on the evolution of haplodiploidy?

I have shown that intuitive arguments correctly predict the effect of inbreeding on the maternal transmission and deleterious mutation hypotheses for the evolution of haplodiploidy. Inbreeding favours the deleterious mutation argument and disfavours the maternal transmission argument. A more difficult question is whether such results tell us whether inbreeding makes the evolution of haplodiploidy more or less likely.

With a single viability locus the selective force of the maternal transmission hypothesis is far greater than the selective force of the deleterious mutation hypothesis. When an *Pp* female mates with an *pp* male, for example, half of the sons are *P* and half *pp*, while half of the daughters are *Pp* and half *pp*. Thus in terms of genetic composition of adults the frequency of *P* has jumped from ¼ in the parents to ½ in the sons and ¼ in the daughters. In contrast, *P* spreads much more slowly by the deleterious mutation hypothesis, at a rate of the order of the mutation rate.

This difference in rates of spread need not apply if multiple viability loci are invoked, as seems biologically realistic (Goldstein, 1994). If the modelling of multiple viability loci has the same effect on the transition from diplonty to haplodiploidy as on the transition from diplonty to haplonty, then the rate of spread of *P* becomes of the order of the per genome mutation rate *U* (Jenkins & Kirkpatrick, 1995). Unless we know the size of *U* it is not possible to know which of the two hypotheses, maternal transmission or deleterious mutations, is likely to be the stronger (see below).

An alternative approach is to compare the sensitivity of the two hypotheses to increasing inbreeding. As can be seen from Table 1, inbreeding changes the critical value of *f* (maternal transmission theory) by considerably more than the critical value of *h* (deleterious mutation theory). Although we do not know the underlying probability distribution of either the *f* or *h* parameters (see below), this result does suggest that inbreeding may restrict the evolution of haplodiploidy, in contrast to previous conclusions (Borgia, 1980).

The conclusions of this study are dependent on the distributions of the parameters *U*, *f* and *h*. What do we know about these parameters?

The deleterious mutation rate per genome *U* has been estimated to be greater than unity in humans (Eyre-Walker & Keightley, 1999) and around unity in *Drosophila* (Simmons & Crow, 1977). These values suggest that the strength of the deleterious mutation effect may be just as strong as the strength of the maternal transmission effect.

Haploid males are likely to be weaker than diploid males for at least four reasons (Whiting, 1945; Borgia, 1980), so we can predict $f < 1$. First, males must develop from unfertilized eggs. Secondly, males may suffer from gene dosage problems. Thirdly, haploid males must be fertile, in

other words spermatogenesis must work. Finally, haploid males will suffer from somatic deleterious mutations more than their diploid counterparts (Orr, 1995). From these considerations we might conclude that the maternal transmission hypothesis is unlikely to work if all matings are between sibs, because then the requirement for the invasion of haplodiploidy is $f > 1$. However, the considerations above need not apply under full inbreeding because haploid males would not enter into competition with diploid males and thus their weakness need not cause a corresponding decrease in fitness (Borgia, 1980).

Penetrance in *Drosophila* depends on the strength of selection (Simmons & Crow, 1977). Lethals are highly recessive ($0.01 < h < 0.03$) while mildly deleterious mutations show greater dominance ($0.3 < h < 0.5$). If genomic recombination is restricted (Goldstein, 1994) and/or if there is heavy inbreeding then the deleterious mutation hypothesis may favour haplodiploidy over diplonty.

This study has been carried out using a fairly simple model of haplodiploidy evolution. Several biological complexities have been ignored, the implications of which may well merit further study.

I have not considered the so-called parahaplodiploidy in which males develop from fertilized eggs but transmit only the female genome (Bull, 1979). The maternal transmission argument still applies to parahaplodiploidy, but the deleterious mutation theory seems unlikely to favour parahaplodiploidy over diplonty.

In the model studied here, I have assumed that the haplodiploidy modifier acts on the whole genome rather than on a single chromosome. However, the evolution of regional haplodiploidy on a single chromosome would probably be little different from the evolution of full haplodiploidy on the entire genome. Selection against deleterious mutations would be weaker for a single chromosome, which would reduce the deleterious mutation effect favouring haplodiploidy. On the other hand, recombination between loci on a single chromosome would be less than recombination between loci spread over the whole genome, which would favour haplodiploidy. There is also the possibility that chromosomes may differ in the penetrance and strength of selection against mutations. With regard to the maternal transmission hypothesis, the spread of a partial haplodiploidy would appear to be no different, in terms of population genetics, from the

spread of full haplodiploidy. However, if only a subset of the genome is spreading selfishly, then perhaps there is greater scope for suppressor genes to evolve.

I have assumed that all mutations are deleterious, despite the fact that advantageous mutations must occur for adaptive evolution to proceed. Favourable mutations will appear at a higher rate in diploid males than in haploid males. If penetrance and recombination are both high then diplonty is favoured over haplonty (Orr & Otto, 1994), and the same conditions probably favour diplonty over haplodiploidy.

In some ways haplodiploidy might be thought of as a intermediate form between diplonty and haplonty, in which case it might not be clear why haplodiploidy is so evolutionary stable. If the transition from diplonty to haplodiploidy is favoured, why not the transition from haplodiploidy to full haplonty? In the case of the deleterious mutation hypothesis, it is true that those conditions which favour haplodiploidy over diplonty will usually favour full haplonty over diplonty (Goldstein, 1994). This conclusion probably holds under inbreeding as well as outbreeding. In the case of the maternal transmission hypothesis it is hard to see how selfish genes could drive the evolution of full haplonty, although a selfish gene process could easily exploit the two-fold cost of sex to spread asexuality. The maintenance of haplodiploidy could then be explained by the instability of asexual lineages. Another possible answer to this problem is that gene dosage problems might be severe. Male haploidy may well be able to evolve despite the heavy costs of haploidy, but the costs of female haploidy may be too high if one considers the greater dependency of population fitness on female rather than male fitness.

Inbreeding raises a number of issues which have been avoided in this study. First, the model I have used does not specify sex determination, but is consistent with XX males and XO females which ensures that haploids are male. Alternative haplodiploid sex determination mechanisms may be costly under inbreeding (Werren, 1993). Secondly, haplodiploidy enables the primary sex ratio adjustment which is often favoured by inbreeding (although inbreeding is neither necessary nor sufficient for selection to favour biased sex ratios (Charnov, 1982)). However, this benefit can only be enjoyed after haplodiploidy has evolved and therefore cannot be considered as an explanation for the evolution of haplodiploidy. Thirdly, intragenomic conflict over the sex ratio might itself drive the evolution of haplodiploidy (Haig, 1993). This explanation is similar to the

maternal transmission hypothesis in that intragenomic conflict is the driving force behind the evolution of genetic systems. Fourthly, another potential advantage of haplodiploidy associated with inbreeding is that haplodiploid mothers can produce sons with which to mate if population density is low (Hartl & Brown, 1970; Borgia, 1980).

One final complication not addressed here is the response to acute inbreeding. In this study I have assumed chronic inbreeding, so that the population reaches the mutation-selection equilibrium under inbreeding before the ploidy modifier is introduced. Alternatively, haplodiploidy may be favoured as a response to acute inbreeding. If inbreeding also favours a female bias to the sex ratio, then haplodiploidy would be especially favoured. Haplodiploidy would enable female fitness, and hence population fitness, to survive the transition from outbreeding to inbreeding relatively unscathed. Males, being haploid, would suffer from the exposure of deleterious mutations far more than the diploid females. But the population would not suffer as long as a few males were fertile. Under this scenario the evolution of sex specific dispersal patterns would come about as a consequence of differences in mutation load. The weak males would be good for little more than inseminating their more vigorous sisters who would then go forth and multiply.

Acknowledgements

Many thanks to Laurence Hurst and two anonymous referees for many useful suggestions.

References

BORGIA, G. 1980. Evolution of haplodiploidy: models for inbred and outbred systems. *Theor. Pop. Biol.*, **17**, 103-128.

BOURKE, A. F. G. AND FRANKS, N. R. 1995. *Social Evolution in Ants,* 1st edn. Princeton University Press, Princeton.

BROWN, S. W. 1964. Automatic frequency response in the evolution of male haploidy and other coccid chromosomal systems. *Genetics*, **49**, 797-817.

BULL, J. J. 1979. An advantage for the evolution of male haploidy and systems with similar genetic transmission. *Heredity*, **43**, 361-381.

CHARNOV, E. 1982. *The Theory of Sex Allocation,* 1st edn. Princeton University Press, Princeton.

EYRE-WALKER, A. AND KEIGHTLEY, P. D. 1999. High genomic deleterious mutation rates in hominids. *Nature*, **397**, 344-347.

GOLDSTEIN, D. B. 1994. Deleterious mutations and the evolution of male haploidy. *Am. Nat.*, **144**, 176-183.

HAIG, D. 1993. The evolution of unusual chromosomal systems in coccoids: extraordinary sex ratios revisited. *J. Evol. Biol.*, **6**, 69-77.

HAMILTON, W. 1993. Inbreeding in Egypt and in this book: a childish view. In: Thornhill, N. W. and Shields, W. M. (eds) *The Natural History of Inbreeding and Outbreeding,* pp. 429-450. University of Chicago Press, Chicago.

HAMILTON, W. D. 1967. Extraordinary sex ratios. *Science*, **156**, 477-488.

HARTL, D. L. AND BROWN, S. W. 1970. The origin of male haploid genetic systems and their expected sex ratio. *Theor. Popul. Biol.*, **1**, 165-190.

HURST, L. D. AND PECK, J. R. 1996. Recent advances in understanding of the evolution and maintenance of sex. *Trends Ecol. Evol.*, **11**, 46-52.

JENKINS, C. D. AND KIRKPATRICK, M. 1995. Deleterious mutation and the evolution of genetic life cycles. *Evolution*, **49**, 512-520.

MABLE, B. K. AND OTTO, S. P. 1998. The evolution of life cycles with haploid and diploid phases. *Bioessays*, **20**, 453-462.

OLIVER, J. H. 1971. Parthenogenesis in mites and ticks (Arachnida: Acari). *Am. Zool.*, **11**, 283-299.

ORR, H. A. 1995. Somatic mutation favors the evolution of diploidy. *Genetics*, **139**, 1441-1447.

ORR, H. A. AND OTTO, S. P. 1994. Does diploidy increase the rate of adaptation? *Genetics*, **136**, 1475-1480.

OTTO, S. P. AND GOLDSTEIN, D. B. 1992. Recombination and the evolution of diploidy. *Genetics*, **131**, 745-751.

OTTO, S. P. AND MARKS, J. C. 1996. Mating systems and the evolutionary transition between haploidy and diploidy. *Biol. J. Linn. Soc.*, **57**, 197-218.

SIMMONS, M. J. AND CROW, J. F. 1977. Mutations affecting fitness in *Drosophila* populations. *Ann. Rev. Genet.*, **11**, 49-78.

VALERO, M., RICHERD, S., PERROT, V. AND DESTOMBE, C. 1992. Evolution of Alternation of Haploid and Diploid Phases in Life-Cycles. *Trends in Ecology & Evolution*, 7, 25-29.

WERREN, J. H. 1993. The evolution of inbreeding in haplodiploid organisms. In: Thornhill, N. W. and Shields, W. M. (eds) *The Natural History of Inbreeding and Outbreeding,* pp. 42-59. University of Chicago Press, Chicago.

WHITING, P. W. 1945. The evolution of male haploidy. *Q. Rev. Biol.*, **20**, 231-260.

Figure Legends

Figure 1: The model for the evolution of haplodiploidy under panmixia.

Figure 2: The conditions for the invasion of diplonty by haplodiploidy in terms of selection against deleterious mutations (*s*) and penetrance (*h*) for different levels of sib mating (*sib*). The minimum value of *h* at which haplodiploidy can invade is given to a precision of 0.001. Goldstein refers to Goldstein's (see equation 5, 1994) analytically derived results for random mating (compare with $sib = 0$).

Figure 3: The minimum haploid fitness relative to diploids (*f*) required for haplodiploidy to spread in a diplontic population with different levels of sib mating (*sib*) under the maternal transmission hypothesis.

Figure 1

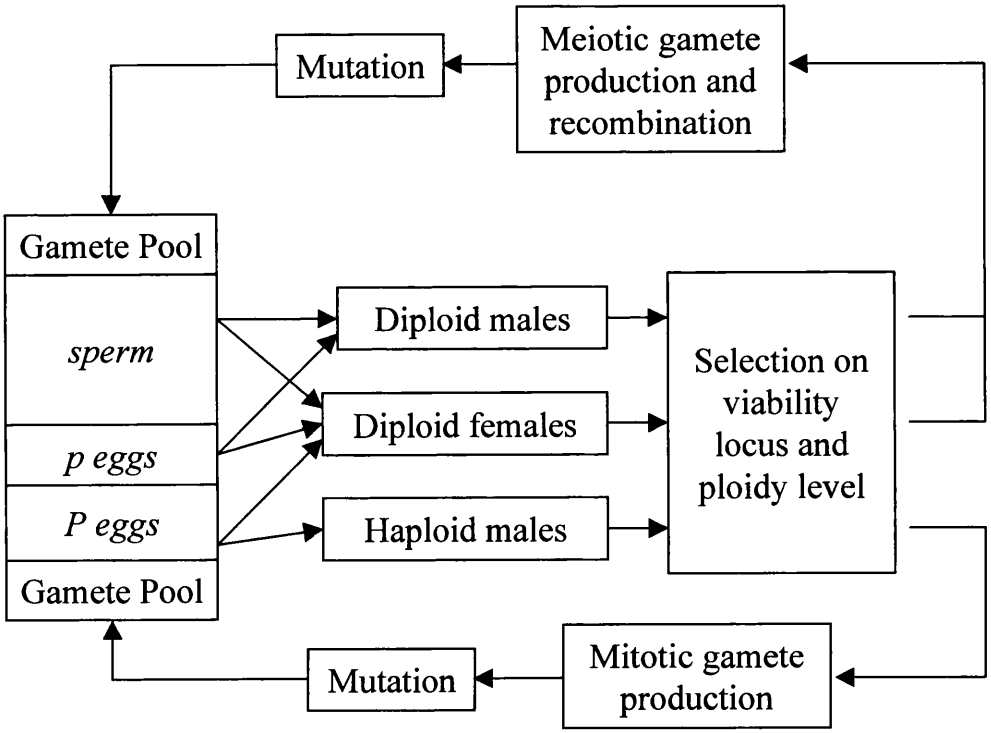


Figure 2

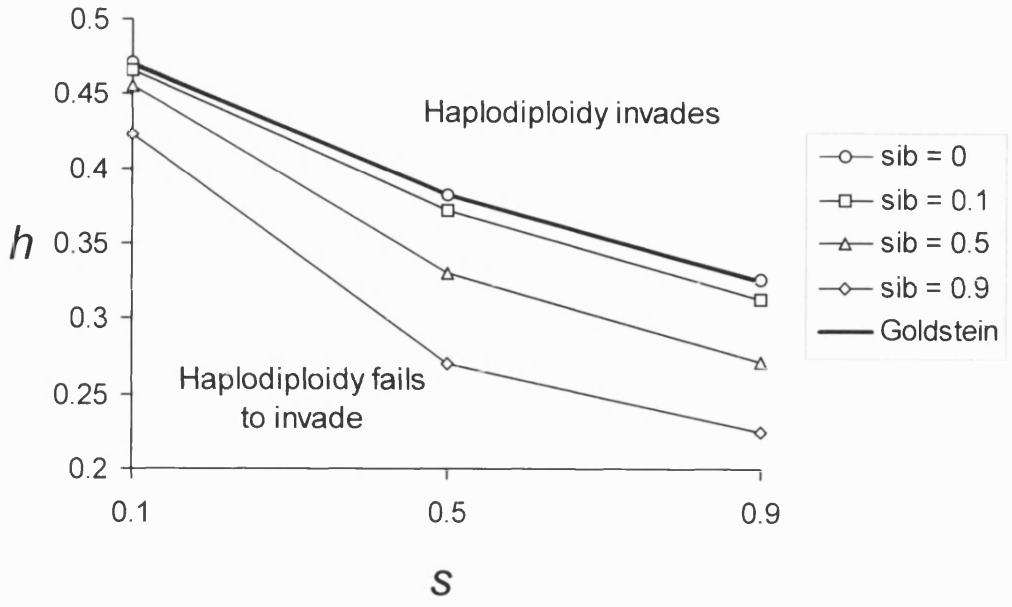


Figure 3

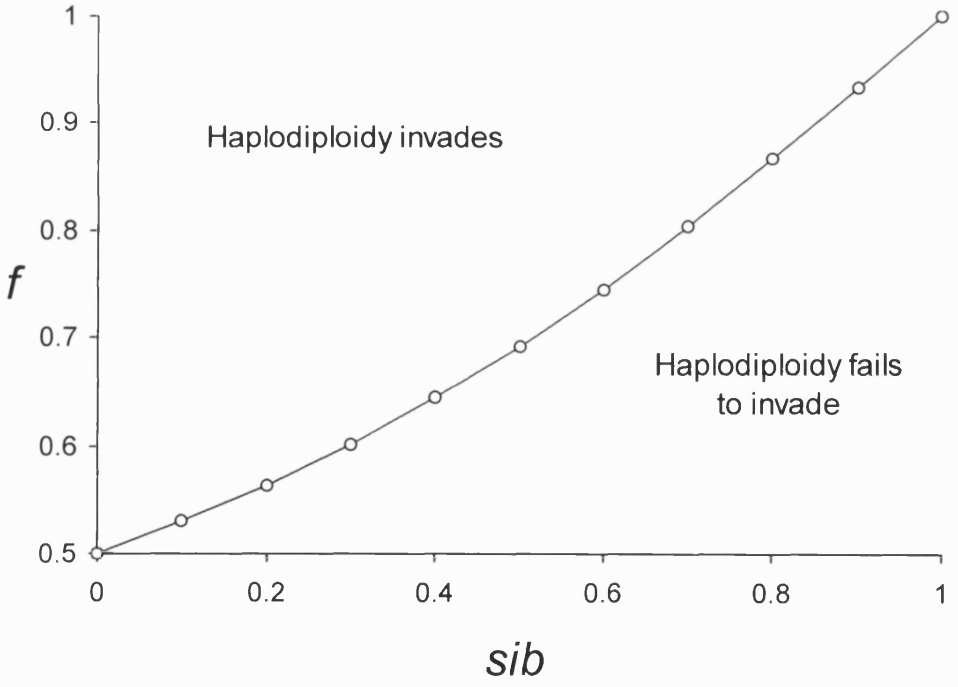


Table 1. The effect of inbreeding on the deleterious mutation and maternal transmission hypotheses for the evolution of haplodiploidy. For different values of selection against deleterious mutations (*s*) and sib mating (*sib*), the minimum values of haploid fitness relative to diploids (*f*) and the minimum values of penetrance (*h*) at which haplodiploidy invades is given. The evaluations of the minimum values of *f* and *h* are to a precision of 0.0001 and 0.001 respectively. Also shown are the percentage changes in *f (%f)* relative to *sib* = 0 as *sib* changes and the percentage changes in *h (%h)* relative to *sib* = 0 as both *sib* and *s* change.

sib	*s*	*f*	*h*	*%f*	*%h*
0	0	0.5	-	0	-
0	0.1	0.5	0.471	0	0
0	0.5	0.5	0.383	0	0
0	0.9	0.5	0.326	0	0
0.1	0	0.5299	-	5.98	-
0.1	0.1	0.5299	0.466	5.98	-1.06
0.1	0.5	0.5299	0.372	5.98	-2.87
0.1	0.9	0.5299	0.314	5.98	-3.68
0.5	0	0.691	-	38.2	-
0.5	0.1	0.691	0.455	38.2	-3.4
0.5	0.5	0.691	0.331	38.2	-13.6
0.5	0.9	0.691	0.272	38.2	-16.6
0.9	0	0.9332	-	86.6	-
0.9	0.1	0.9332	0.423	86.6	-10.2
0.9	0.5	0.9332	0.271	86.6	-29.2
0.9	0.9	0.9332	0.225	86.6	-31.0

Chapter 3. The evolution of polyploidy

In Chapter 2 I considered the relative merits of haplonty, diplonty, haploid-diploidy and haplodiploidy. In this chapter I shall consider the evolution of higher ploidy copy numbers. It seems reasonable to suppose that the broad conclusions of the genetic hypotheses can be extended from the haploid-diploid transition to the diploid-tetraploid transition: ploidy level increases are favoured by (i) high recombination, (ii) low penetrance of deleterious mutations, (iii) high penetrance of favourable mutations, and (iv) synergistic epistasis or truncation-like selection. It also appears that the ecological arguments given above concerning ploidy levels should apply to polyploids just as they do to haploids and diploids.

Polyploidy in asexual organisms

As mentioned above, polyploid asexual eukaryotes do exist (Kondrashov 1994; Raikov 1982), for example the protozoan *Giardia lamblia* appears to possess 3 to 5 identical or nearly identical copies of its haploid genome (Fan *et al.* 1991). Such polyploidy might be favoured by environmental factors, as Kondrashov (1994) has suggested. Increases in ploidy appear to occur by endomitosis and decreases in ploidy by reduction (Hurst and Nurse 1991; Kondrashov 1994). The nature of the asexual ploidy cycle thus permits any level of ploidy, in marked contrast to the case for sexual organisms.

Polyploidy in sexual organisms

In unicellular organisms no distinction can be drawn between germline and somatic ploidy levels. In multicellular organisms, however, somatic ploidy levels need not equal germline ploidy levels. The following summary of somatic ploidy levels has been gleaned from the book by Brodsky and Uryvaeva (1985).

Somatic ploidy levels can increase by polyploidy, the doubling of chromosome number by endomitosis, or by polyteny, the repeated doubling of the number of chromatids in the chromosomes. Somatic polyploidy has been reported in mammals, insects and plants. High ploidy cells are often terminally differentiated, and it has been suggested that polyploidisation is a way of attaining differentiation. In keeping with the general relationship between ploidy and size, polyploidy increases both cell and nucleus size, although the consequent reduction in the surface area to volume ratio is often countered by various shape changes. The large size of polyploid cells appears to be functional in a number of cases: the epithelium of the urinary bladder, butterfly wing scale cells, and the ciliated cells of the starfish stomach. Polyploid cells are more resistant than diploid cells to both damage and mutation of DNA, although polyploid cells are also associated with carcinogenesis.

Germline polyploidy appears to be less common than somatic polyploidy. One restriction on polyploidy in sexual organisms is due to chromosome segregation in meiosis: odd levels of ploidy lead to the production of unbalanced gametes (Futuyma 1998). The mechanism of meiosis means that although a genome may be polyploid, it is forced to behave as though it were diploid. This effective diploidy explains how it is possible to claim that we humans are both diploids (from our life cycle) and octoploids (since the vertebrates may have evolved by two polyploidisation events). The fact that many duplicated genes either evolve new functions or undergo mutational decay (see Chapter 4) also means that polyploidy tends to decay into diploidy. Germline polyploidy is thought to have been important over evolutionary history, because it allows immediate speciation (Futuyma 1998; Haldane 1932) and may open up new developmental possibilities.

Polyploidy is common in plants, with about 70% of angiosperms appearing to have polyploidy in their ancestry (Masterson 1994). For example, the polymodal distribution of genome sizes in the monocotyledons suggests a series of genome duplications (Sparrow and Nauman 1976). Experimental results suggest that the genome multiplication *per se* is selectively disadvantageous, but that allotetraploidisation can be favoured by the increase in genetic diversity (Li 1997). Polyploidy is much rarer in animals (Muller 1925; Orr 1990), except for a few groups such as fish and the anuran amphibians (Ptacek *et al.* 1994). This difference is thought to be due to the fact that Y chromosome degeneracy is more common in animals than plants (Orr 1990). The degeneration of the Y chromosome leads to dosage compensation and since polyploidy destroys the balance of dosage compensation systems polyploidy is selectively disadvantageous in most animals. Those animal groups which display polyploidy appear to have weakly differentiated sex chromosomes (Ohno 1970).

Polyploidy in vertebrate evolution

Maynard Smith and Szathmary (1995) have considered the major transitions in evolution in terms of the acquisition of news levels of organisation and selection. Holland (1998) has proposed an alternative view of the major transitions in animal evolution by emphasising changes in developmental complexity. Two of the major transitions identified by Holland, the origin of the vertebrates and the origin of the gnathostomes, are thought to have arisen by genome duplications, that is by polyploidisation events. Here I focus on the role of polyploidy in vertebrate evolution, first proposed by Ohno (1970). It is thought that whereas single gene duplications allow only small changes in development, genome duplications allow "saltational increases in developmental complexity" (Sidow 1996).

Following the duplication event, and its spread through the population, redundancy will probably lead to high rates of sequence evolution in most duplicated genes (see Chapter 4). Mutational decay can be halted by one of the copies evolving a new function, in developmental terms when it is co-opted into a novel regulatory framework (Sidow 1996). Such a process may be

aided by the existence of "junk" DNA caused by polyploidisations which increases the likelihood of mutations to the enhancers of master control genes (Ruddle 1997; Sidow 1996). The duplication of entire regulatory pathways means that new developmental processes can evolve more easily following polyploidisation (Holland and GarciaFernandez 1996). The metabolic costs of this genome expansion may have been offset to some extent by the evolution of methylation-based gene repression mechanisms (Bird 1995).

Spring (1997) has suggested that the vertebrate evolution may have proceeded by allotetraploidy rather than autotetraploidy. With allotetraploidy the initial level of redundancy in the tetraploid is reduced, which means that duplicated genes are more likely to be retained and evolve new functions. As long as the two species have not diverged so much that the initial tetraploid is too weak, allotetraploidy should enable more developmental possibilities to be explored than autotetraploidy (Spring 1997).

We know that genome duplications have occurred in evolutionary history (for evidence of an ancient yeast duplication see Wolfe and Shields 1997), but can genome duplications explain the evolution of vertebrate genomic anatomy? The most commonly held model of vertebrate evolution posits two separate tetraploidisation events about 500 and 430 MYA (Sharman and Holland 1996; Skrabanek and Wolfe 1998). Such a model is consistent with, but certainly not proved by, current data on gene numbers, tetralogies and homologous clusters (Skrabanek and Wolfe 1998). As discussed further in Research Paper 3, it is difficult to test hypotheses of ancient duplications, the traces of which may have been obscured by subsequent events.

The *Hox* genes provide evidence for the ways in which gene duplications can lead to increases in developmental complexity, as well as providing useful tests of the polyploid theory (Finnerty and Martindale 1998; Holland and GarciaFernandez 1996; Ruddle 1997). *Hox* genes are important in development because they encode positional identity along the anteroposterior body axis. The evolution of vertebrate developmental mechanisms can be analysed by comparative methods, which require *Hox* gene data and reliable chordate phylogenies (Finnerty and Martindale 1998; Holland and GarciaFernandez 1996). Can *Hox* gene evolution provide a genetic basis for the possession in vertebrates of numerous morphological characters, mostly in the head region, not found in other chordates? It does appear that "the most extensive phase of elaboration of the chordate *Hox* clusters correlates with the most extensive phase of developmental elaboration and innovation, close to vertebrate origins" (Holland and GarciaFernandez 1996). However, the evolution of *Hox* clusters is turning out to be rather complex, with loss and gain of single genes, cluster duplications, cluster splitting and partial cluster inversions (Finnerty and Martindale 1998; Ruddle *et al.* 1994). Indeed there appears to be "no simple relationship between *Hox* gene evolution and morphological evolution" (Finnerty and Martindale 1998).

There appears to be rather more to vertebrate evolution than just two polyploidisation events. A comparison of the regions containing the *Hox* clusters in mouse and human has revealed that the four *Hox* clusters evolved by three, rather than two, block duplications (Bailey *et al.* 1997).

The study also revealed evidence in favour of a "semiconservative model" of duplication, whereby one duplication product retains the functional attributes of the precursor (two clusters are involved with neural tube and somite pattern formation) while the other duplication product takes on new characteristics (two clusters are involved with vertebrate specific limb and urogenital pattern formation) (Ruddle 1997). In Chapter 4 I shall discuss the evolution of such homologous clusters in vertebrates and investigate whether homologous clusters are the result of genome duplications or single gene duplications followed by rearrangements.

Chapter 4. The evolution of multiple gene copies

What are genes?

Before describing the evolution of genomic anatomy at the level of the gene, I should first define what I mean by the term *gene* since different groups of biologists use the same term in different ways.

In theoretical population genetics a gene is any heritable factor which shows concomitant variation with phenotype, as in "the gene for baldness". This approach is general and can be applied to any system of heredity (Frank 1996). The population genetics definition of the gene has been narrowed by Dawkins (1976) to include only those heritable units upon which selection can act. In contrast, the practical approach of molecular biologists is to define a gene as a nucleotide sequence which encodes a product, as in "the gene for the insulin-like type II growth factor receptor".

The population genetics notion of the gene is rather complicated by pleiotropy and epistasis. If multifunctionality and overlapping redundancies really are common then the population genetics definition becomes unwieldy (Wagner 1998a). Complex developmental systems mean that the relationship is not between genotype and phenotype, but between genotypes and phenotypes. It becomes impossible to name a gene on the basis of its one phenotypic effect as in "the gene for breast cancer". Long ago Ernst Mayr (1955; 1963) made similar points based on the view that coadapted gene complexes were of great importance. The population genetics approach which Mayr derided as "bean bag genetics" was strongly defended by Haldane (1964), so we now find that "The modern view, spurred on by the rush of DNA data, is that we can profitably study loci in isolation" (Gillespie 1998).

There is little reason to suppose that there should be a single definition of the gene. Different groups of biologists use those definitions which help them. But in the context of genomic anatomy an explicit nucleotide sequence approach is clearly useful. I shall use the term gene to mean modular transcriptionally active units, such as protein-coding genes and RNA-coding genes. Thus the genome consists of both genic and nongenic DNA. This is the standard usage of molecular evolution (as in Li 1997).

Having defined what genes are, what features of the genomic anatomy of genes need to be explained? I shall focus one three issues. In Chapter 4 I shall consider the evolution of multiple copies of genes. In Chapter 5 I discuss the causes and consequences of monoallelic expression. In Chapter 6 I examine the issue of why different genes evolve at different rates.

The evolution of complex genomes

How does a genome gain functional complexity? There appear to be two alternatives: either increase the functional complexity of existing genes or increase the number of genes.

The former strategy can be accomplished by a number of different mechanisms. Internal duplication of functional domains allows the gene to take up additional functions or to improve its current function, and many complex genes appear to have arisen by internal duplications which cause internal repeats (Li 1997). The presence of 15 internal repeats within the mammalian *Igf2r* gene (Morgan *et al.* 1987) implies that it acquired its current functional complexity by domain duplication. Alternative ways of increasing functional complexity without increasing gene copy number include alternative splicing (Smith *et al.* 1989), overlapping genes (Li 1997), gene sharing (Piatigorsky and Wistow 1989), and RNA editing (Chan 1993). The evolution of multiple functions is thought to have been common in the evolution of developmental control genes which can gain new functions by the acquisition of enhancers, as opposed to terminal genes which have to yield a functional protein product.

Ohno's (1970) assertion that gene duplication is the only means of gaining new function is falsified by the above mechanisms, but gene duplication is widely held to be the most important process in the gain of new functions, as demonstrated by the existence of gene families (Li 1997). In Chapter 3 I discussed the evolution of polyploidy by duplication of the entire genome and in this chapter I shall focus on processes on duplication at the level of genes or groups of genes. It seems likely that gene duplication events either take place by polyploidisation, in which case the dosage of all genes is increased together, or as gene duplications in which the dosage of only one gene is increased. Duplications of intermediate size such as chromosomal duplications would probably cause dosage compensation problems, as demonstrated by the negative fitness consequences of Down's syndrome in humans and partial polysomy in *Drosophila* (Li 1997).

Gene duplication

The duplication of a gene generates two copies. As long as the entire gene is copied both duplicates should be initially functional. What then happens?

The neutralist view of gene duplications suggests that one of the duplicates either gains a new function or becomes a pseudogene. This argument is based on the assumption that two copies per haploid genome are no better than one, and thus the duplication mutant must have spread through the population by random drift. In Ohno's (1973) hypothesis one of the duplicates would be free to evolve neutrally, usually becoming a inactive pseudogene but sometimes evolving a new function: "duplications enable genes to make evolutionary experiments which have been previously forbidden, liberating them from the incessant natural selection whose overwhelming activity is

eliminating variants" (Kimura 1983). The neutral evolution of the duplicate might produce a new gene which then becomes favoured under new environmental conditions, an example of what Kimura (1983) termed the "Dykhuizen-Hartl effect".

As an alternative to the classical neutral evolution model, one might imagine that selection might affect duplicated genes. Purifying selection does appear to act on duplicates (Hughes and Hughes 1993) and positive selection seems to drive the evolution of new roles (Hughes and Nei 1988). The effect of positive selection would favour the initial spread of the duplication through the population and favour the acquisition of new roles. Negative selection might also favour the preservation of duplicates which become specialists at the multiple roles performed by the original gene (Force *et al.* 1999; Hughes 1994).

The mutation rate of repeated sequences may be high because of cosuppression, the process of methylation and mutagenesis which acts on duplicated sequences (Garrick *et al.* 1998; Jensen *et al.* 1999; Kricker *et al.* 1992). Cosuppression only works for duplicates of high sequence similarity (Kricker *et al.* 1992), and so will only drive diversification up to a certain point. The role of cosuppression may be to act against the horizontal spread of selfish genes within the genome.

So unless one of the copies gains a new advantageous function the steady accumulation of mostly deleterious mutations will eventually lead to one copy becoming functionless (Kimura and King 1979). Because deleterious mutations are thought to be more likely than advantageous ones, theoretical studies have predicted that rates of gene loss should be at least an order of magnitude higher than rates of gene divergence (Wagner 1998a). Using the proposed polyploidisation events of vertebrate evolution (see Chapter 3), Nadeau and Sankoff (1997) have shown that the rate of gene divergence is about the same as the rate of gene loss. The discrepancy between theoretical population genetics and the empirical data may be related to the existence of multifunctional proteins since multifunctionality implies that the acquisition of new roles may be relatively easy (Wagner 1998a). For the duplicate to survive it need not devote itself entirely to a completely new role. Cooke *et al.* (1997) have suggested that a duplicate can be retain its original function if it acquires an extra role. Thus evolution by gene duplication is intimately related to the multifunctionality of genes and overlapping redundancies. Gibson and Spring (1998) have questioned Cooke *et al.*'s (1997) arguments for the adaptive evolution of redundancy and have instead suggested that redundancy is found because the mutational decay of multidomain duplicates proceeds slowly since many deleterious mutations will cause dominant negative phenotypes. There is an alternative way in which multifunctionality might aid the retention of duplicates: if the original gene is multifunctional the two copies may become single function genes specialising in different functions (Force *et al.* 1999; Hughes 1994).

The assumption that two identical functional copies are no better than one may not hold for a number of reasons. First, a single locus may not be able to produce gene product at a fast enough rate. This dosage requirement appears to explain multiple copies of genes for rRNAs and tRNAs

(Li 1983), histones (Elgin and Weintraub 1975), and testis-specific Y-linked genes (Lahn and Page 1997). The effect of gene dosage may be responsible for the arms race which has lead to high copy number of *Stellate* in *D. melanogaster* (Hurst 1996). Secondly, multiple genes may protect against the effects of stochasticity at low molecular concentrations (McAdams and Arkin 1999). The complexity and sensitivity to stochastic noise of developmental pathways may mean that duplicates are more likely to be useful for development control genes than "housekeeping" genes. This is the developmental equivalent to the somatic mutation argument for diplonty over haplonty (see Chapter 2). Thirdly, multiple genes provide buffering against the effect of germline mutations. But as with ploidy levels, multiple gene copies should cause an eventual increase in load (see Chapter 2). Thus Clark (1994) found in mutation-selection-duplication model that duplications *per se* have to be beneficial to spread. However, Wagner (1999) has shown that natural selection can favour an increase in redundancy, though this result is probably dependent on his assumption of complete recessivity, the supposition that redundancy provides complete buffering. When deleterious mutations are completely recessive then the mutational load is the same in diplonts and haplonts (Crow and Kimura 1970).

Genome size

The genome size of an organism is defined by the size of its haploid genome, its "characteristic" size. Hence the C value paradox refers to the problem of genome size variation. The C value paradox is particularly intriguing for the eukaryotes which display an 80,000 fold range of C values. Evidence suggests that this genome size variation is not due to variation in the number of single copy sequences. There appears to be no relationship between variation in genome size and either organismal complexity (Nei 1969), evolutionary specialisation (Britten and Davidson 1971), or the predicted number of different genes (Cavalier-Smith 1985). Furthermore, the reassociation kinetics of eukaryotic DNA demonstrate that much of the genome consists of repetitive sequences (Lewin 1997).

So sequence copy number variation is required to explain genome size variation. What kinds of sequences are present in multiple copies and does such copy number variation make evolutionary sense? The rRNA genes are found in multiple copies in eukaryotes, and indeed copy number appears to correlate with genome size (Li 1997). In addition many genes belong to gene families and superfamilies, although such groups contain variant sequences. But such sequences constitute only a very small proportion of the genome, and so to answer the C value paradox it is necessary to consider the copy number variation of nongenic DNA and transposable elements (Cavalier-Smith 1985).

Transposable elements, capable of horizontal spread through the genome, constitute much of the moderately dispersed repeated DNA of eukaryotes (Berg and Howe 1989). Transposable elements are particularly common in heterochromatin, but the implications of this trend with

respect to the evolutionary forces acting on transposable elements remain unclear (Dimitri and Junakovic 1999).

How can one make sense of the fact that the bulk of the genome is composed of repeated units? There are three classes of explanation which differ with respect to the role of selection on the repeated units. The functional explanation suggests that the repeated units are favoured by selection at the genomic level, either for genetic (Zuckerkandl 1976) or chemical (Hsu 1975; Vinogradov 1998) or physical properties (Cavalier-Smith 1978; Zuckerkandl and Hennig 1995). The junk explanation proposes that the repeated units are unaffected by selection at the genomic level (Ohno 1972). The conflict explanation proposes that the spread of repeated units is advantageous at the level of the repeat units themselves but is deleterious at the level of the genome (Doolittle and Sapienza 1980; Orgel and Crick 1980; Ostergren 1945).

The functional explanation suggests that multiple copy numbers make sense in terms of selection on individuals. Some nongenic repetitive DNA, such as telomeric sequences (Zakian 1989) and the *Responder* satellite sequence in *D. melanogaster* (Wu and Hammer 1991), appears to be functional, although such properties of heterochromatin may not explain the origins of their constituent repetitive elements (Charlesworth *et al.* 1994). The functional explanation is supported by little more than circumstantial evidence, and the predicted deleterious effects of the deletion of nongenic DNA have not been observed (John and Miklos 1988). Zuckerkandl (1976) has suggested that the noncoding bulk of the DNA helps to globally regulate gene expression, and Zuckerkandl and Hennig (1995) have suggested a role for repetitive elements in regulating chromatin structure. Both these theories are compatible with the finding that repetitive DNA causes changes in transcriptional activity and chromatin structure (Garrick *et al.* 1998), but a non-functional view of such phenomena as an adaptive response to selfish DNA appears more likely (Wolffe 1998). Cavalier-Smith (1978) has suggested that genome size is selected via its effects on growth strategy as a result of the links between genome size, nuclear size, cell size and organism size. Cavalier-Smith's theory appears to suppose no alternative way of regulating growth, which appears unlikely (Arendt 1997). Hsu (1975) has suggested that repetitive DNA protects the functional DNA from mutations, and Vinogradov (1998) has proposed that repetitive DNA acts as a buffer against changes in chemical solute concentrations.

The junk hypothesis suggests that much of the repetitive DNA is neutral, or nearly neutral (Ohno 1972), and hence that the existence of multiple copies makes sense in terms of deterministic patterns of mutation generating duplicates. Most nongenic tandem repeats are good candidates for being junk, since such elements are transient over evolutionary time (Li 1997). Although large amounts of repetitive DNA may engender some metabolic and mutational costs (Li 1997), selection in mammals may be too weak to act against tandem repeats. The greater efficacy of selection against genome expansion in *Drosophila* due to greater effective population sizes is indicated by the small size of *Drosophila* introns (Mount *et al.* 1992). An alternative mutational explanation is that *Drosophila* has a high rate of deletions (Petrov and Hartl 1998; Petrov *et al.* 1996). Support

for the junk hypothesis has been claimed from comparative analysis of salamanders (Pagel and Johnstone 1992), although it has suggested that the observation of a correlation between genome size and developmental rate is consistent with both the junk and functional hypotheses (Jockusch 1997).

The conflict explanation supposes that multiple copies make sense in terms of intragenomic selection. Transposable elements are good candidates for the conflict hypothesis of genome size. The act of duplicative transposition spreads the sequences around the genome, and the mutations caused by transposition will usually be deleterious (Charlesworth *et al.* 1994). The possible selfishness of tandem arrays is indicated by human diseases caused by trinucleotide repeats. The enormous amount of repetitive DNA must entail some sort of metabolic cost, though methylation mediated silencing may reduce this effect (Bird 1995).

Concerted evolution

Concerted evolution acts to homogenise repeated sequences and arises as a consequence of the various molecular mechanisms of turnover within the genome (reviewed by Elder and Turner 1996; term coined by Zimmer *et al.* 1980). The effect of concerted evolution on a gene family is to increase similarity within species and to decrease similarity between species. Thus concerted evolution makes evolutionary sense in that it occurs as a result of deterministic mutational processes. Concerted evolution is related to all three aspects of the genomic anatomy of genes (see Chapter 1): identity, copy number and arrangement. Concerted evolution acts on multiple copies, and the rate of concerted evolution is dependent on the levels of identity and the arrangement of repeats. The rate at which concerted evolution proceeds depends on the levels of identity within the gene family. This appears due to the fact that the various processes thought to cause concerted evolution – gene conversion, unequal crossing over, and replication slippage – depend on sequence similarity. Thus gene families whose members diverge rapidly to take on new roles will be less affected by concerted evolution than gene families such as the rRNAs whose members remain very similar. Concerted evolution also acts faster on gene families which are in tandem arrays.

Apart from the mutational explanation for concerted evolution there is also a mutually compatible selective explanation. If the process of concerted evolution is selectively advantageous, then selection favours those mutational forces which are responsible for concerted evolution.

What are the evolutionary implications of concerted evolution? Concerted evolution may increase the efficacy of selection because it enables selection to act on many genes at once (Arnheim 1983). However, Dover (1982) has suggested that concerted evolution may override selection by the process of molecular drive. The rate of homogenisation within the family is so much greater than the rate of appearance of new mutations that there is little variation upon which

selection can act: "the individuals might be turning into monsters, yet there would be no significant increase in variance in the population in the degree of monstrosity" (Dover 1982).

Multiple copies will be identical immediately following duplication, and so concerted evolution will act to slow the divergence of duplicates. Concerted evolution may be disadvantageous if it prevents the acquisition of new roles, but conversely concerted evolution may prevent mutational drift to non-functionality. Concerted evolution may also affect levels of genetic variation (see Li 1997).

I have studied the evolution of concerted evolution with respect to deleterious mutations. If concerted evolution could always choose a wildtype copy to convert the other copies then clearly concerted evolution would be favoured. But if concerted evolution could only choose loci at random, then would homogenisation still be favourable? Ohta (1989) has shown that homogenisation is always favourable in terms of mutational load, since homogenisation means effective haploidy, and haploid load is lower than diploid load. But load arguments only apply to isolated populations. To discover whether homogenisation can evolve within a population I used modifier analysis (see Research Paper 2).

I found that homogenisation is favoured under the same conditions which favour the evolution of haplonty over diplonty (as in Otto and Goldstein 1992). Thus homogenisation is favoured when genes are dominant, in other words if fitness is strongly dosage dependent. Dosage dependency is also a condition which favours the initial evolution of multiple invariant copies, and thus concerted evolution is more likely to be favoured if multiple copies have evolved. Recombination acts against haploidy/homogenisation and so concerted evolution is more likely to be favourable for linked tandem arrays. Concerted evolution proceeds faster within tandem arrays, and so concerted evolution is more likely to be favoured in those cases where it acts fastest. Finally, the strength of selection in favours of modifiers of concerted evolution is only of the order of the mutation rate. So unless mutation rates are very high or modifiers can affect several arrays, population sizes must be large for modifiers of concerted evolution to spread.

This study of homogenisation suggests that features of genomic anatomy affected by concerted evolution may well be adaptive. Those genes which are dosage sensitive, in other words with dominant mutations, will be present in homogenised linked multiple copies. Their effective haploid asexuality will minimise their mutational load. Those genes which are not dosage sensitive, which have recessive and synergistic mutations, will be present in single recombining copies. Such genes will be effectively diploid and sexual, the conditions under which their mutational load will be minimised according to the mutational deterministic theory (Kondrashov 1988). Thus with respect to both genetic system and genomic anatomy, the genome appears to "have its cake and eat it".

Gene arrangement

What factors influence the location of genes in the genome? What chromosome should a gene be on? And how should it be arranged with respect to other genes? Although variation for gene arrangement may well be limiting (Hurst 1999), it is worth attempting to formulate some selective hypotheses of gene arrangement, if only to reject them.

All autosomes are transmitted in the same way from one generation to the next, but sex chromosomes differ in their transmission through the sexes, which means that genes which have sex specific fitness effects will favour a particular sex chromosome (Hastings 1994; Rice 1984). Sexually antagonistic genes can afford to be deleterious to the sex through which they are not transmitted (as demonstrated by Rice 1992). There is some evidence to suggest that some genes on the Y chromosome are sexually antagonistic (Schafer 1994) and many are testis-specific (Lahn and Page 1997). The situation on the X chromosome is less clear, but evidence is now emerging of differences between the X chromosome and the autosomes in gene composition. There is an apparent excess of sex- and reproduction-related genes on the X chromosome as one would expect from considerations of sexually antagonistic and sexually selected genes (Reinhold 1998; Saifi and Chandra 1999). Such clustering of gene types are not confined to the sex chromosomes, with several chromosomes significantly richer than others in both muscle (Pallavicini *et al.* 1997) and extraembryonic expressed (Ko *et al.* 1998) genes. The problem with an anecdotal approach is that "we have no good indication of the proportion of genes on each chromosomal type that are optimally suited to being where they are" (Hurst 1999), and therefore large scale studies are necessary.

How can we explain the arrangement of genes relative to each other, in other words the evolution of linkage and gene order? There are three classes of explanation: history and chance; effects of physical association; and effects of linkage disequilibrium.

History and chance: The existence of gene clusters may reflect the processes of gene origins (the "natal" model of Lawrence and Roth 1996) and gene rearrangements. For example, tandem duplications will yield a cluster of genes with related functions (Horowitz 1945). If subsequent gene rearrangements are selectively neutral then the end product of the shuffling of linkage and gene order by such processes as inversion, duplication, deletion, transposition, and recombination will be due to chance. Under this model the persistence of gene clusters is because mutation is limiting, rather than because of any selective advantage. It is important not to make the assumption that the presence of a feature is a demonstration of its utility, although early studies on bacterial operons fell into this trap: "the mere existence of such arrangements shows that they must be beneficial, conferring an evolutionary advantage on individuals and populations which exhibit them" (Demerec and Hartman 1959).

The "natal" model appears to hold for some eukaryotic gene family clusters, but cannot explain the clustering of genes which are not paralogous, although such patterns could be the result of random shuffling. Historical remnants from bacteria do not seem to be an adequate explanation of gene clusters in either fungi or *C. elegans*, since the gene clusters in these eukaryotes do not seem to have prokaryotic counterparts (Blumenthal 1998; Keller and Hohn 1997).

Random chance may well explain the clustering of genes whose functions are not obviously related (Huxley and Fried 1990), although this is an argument based on ignorance. The risks of making assumptions of non-functionality is demonstrated by the β globin cluster in mammals, which has been claimed to support the natal hypothesis (Maniatis *et al.* 1980). However, it has been recently shown that the regulation of the cluster is dependent on both gene order and orientation (Tanimoto *et al.* 1999), and thus the β globin cluster in mammals now supports the physical association model, which I now consider.

Physical association: Gene clusters can be efficiently coregulated by control at a single site, as appears to explain bacterial operons (Jacob and Monod 1962), although clustering is not a prerequisite for coregulation (Lawrence and Roth 1996). In eukaryotes, polycistronic transcription is also found, with clustered genes not only functionally related but also cotranscribed from a single promoter (Blumenthal 1998). The best examples of eukaryotic polytranscription units come from *C. elegans* in which up to a quarter of all genes are arranged in operons (Blumenthal 1995), though examples have been found in both mouse (Lee 1991) and *Drosophila* (Brogna and Ashburner 1997). Fungi provide strong evidence in favour of the coregulation hypothesis, with many but not all dispensable metabolic functions organised into gene clusters (Keller and Hohn 1997).

The conservation of gene order, not just physical proximity, in bacterial genomes provides additional evidence of coregulation (Dandekar *et al.* 1998). In the majority of such cases the protein products of the linked genes have been shown to physically interact (Dandekar *et al.* 1998). The possibility of cotranslation folding provides a regulatory explanation, although coadaptation (see below) could also favour clustering.

An alternative argument for close physical association of genes is genomic efficiency, with selection favouring deletions of intervening DNA (Blumenthal 1998). Physical association is also favoured by the "selfish operon" model, which suggests that intermittently expressed operons exist so as to favour their wholesale horizontal transmission (Lawrence and Roth 1996). The advantage of physical linkage to horizontal transmission has also been raised in the context of fungal gene clusters (Blumenthal 1998).

Linkage disequilibrium: The evolutionary advantages of sex and recombination are equivalent to the benefits of breaking up linkage disequilibria (Hurst 1999). As described above, deleterious mutations cause recombination to be favoured between loci with synergistic epistasis (Kondrashov 1988), whereas linkage is favoured for multiple loci with antagonistic epistasis.

If specific alleles work well together, if there is coadaptation of genes, then selection can favour the physical clustering of genes (Bodmer and Parsons 1962; Fisher 1930). Examples of coadaptation are provided by meiotic drive genes which usually consist of a toxin linked to an antidote (Lyttle 1991), and the supergene complexes controlling heterostyly in *Primula* and Batesian mimicry in butterflies (Korol *et al.* 1994).

All three explanations for gene clustering probably contain some element of truth, and so the problem becomes one of determining which factor is most important in each case. The greatest problem appears to be a lack of knowledge of mutational biases: how often do paralogs appear and to what extent do genes move around the genome? I have examined these questions in detail with regard to the evolution of homologous gene clusters (see Research Paper 3).

A set of homologous clusters is defined on the basis of each cluster containing members of the same gene families (Lundin 1993). How do homologous clusters evolve? There are three classes of explanation, related to the explanations considered above for gene clustering.

The historical explanation suggests that homologous clusters are generated by block duplications of an initial cluster. The random mutation explanation proposes that homologous clusters will occasionally be created by random gene rearrangements. Finally, the adaptive explanation is that homologous clusters are selectively favoured, because of the benefits of either physical linkage or linkage disequilibrium. One set of homologous clusters on human chromosomes 6, 19, 1 and 9 seems unlikely to have arisen by chance, and has been used to discriminate between the alternative hypotheses of block duplication and adaptive clustering (Endo *et al.* 1997; Hughes 1998; Katsanis *et al.* 1996).

The discrimination of these alternative hypotheses is not easy. Hughes (1998) has suggested one way in which the block duplication hypothesis can be tested by molecular phylogenetics, and has claimed that block duplication cannot explain the data. However, Hughes' reasoning assumes that no gene loss has occurred and thus ignores the possibility of paralogy misallocation. I have shown that Hughes' data are consistent with block duplication if one allows the possibility of gene loss (see Research Paper 3), which has been shown to be roughly equal to the rate of gene divergence (Nadeau and Sankoff 1997). If ones accepts the existence of gene loss then it becomes difficult to discriminate between the alternative hypotheses of block duplication and adaptive clustering. In addition the two explanations are not mutually exclusive since the initial cluster may have arisen for selective reasons, then duplicated, and then the resultant homologous clusters may have been selectively maintained (Lars Lundin, pers. comm.). There is also the problem that phylogenetic reconstruction is often unreliable, especially for inferring ancient evolutionary events (Philippe and Laurent 1998).

The acquisition of more sequence data might provide ways round some of the problems. The issue of gene loss might be resolved in favour of the block duplication hypothesis by the existence of tandem duplicates in early diverging chordates (see Research Paper 3 for details).

More detailed mapping information would allow the use of potentially discriminatory gene order information. Analysis of intronic content within genes could indicate the effects of retrotransposition, a means of gene rearrangement consistent with the adaptive hypothesis. The analysis of numbers of homologous clusters would be informative with respect to the nature of possible block duplications: if all homologous clusters were present in sets of four then two rounds of polyploidisation would be favoured. Finally and most generally, the analysis of large amounts of sequence and mapping data will allow the estimation of mutational biases and thus allow likelihood tests of different hypotheses.

Research Paper 2. The evolution of concerted evolution

Laurence Hurst and Nick Smith (1998)
Proc. Roy. Soc., Lond B **265** 121-127.

The Royal Society

The evolution of concerted evolution

Laurence D. Hurst* **and Nicholas G. C. Smith**

Department of Biology and Biochemistry, Centre for Mathematical Biology, University of Bath, Claverton Down, Bath BA2 7AY, UK

Concerted evolution is a consequence of processes that convert copies of a gene in a multigene family into the same copy. Here we ask whether this homogenization may be adaptive. Analysis of a modifier of homogenization reveals (1) that the trait is most likely to spread if interactions between deleterious mutations are not strongly synergistic; (2) that selection on the modifier is of the order of the mutation rate, hence the modifier is most likely to be favoured by selection when the species has a large effective population size and/or if the modifier affects many genes simultaneously; and (3) that linkage between the genes in the family, and between these genes and the modifier, makes invasion of the modifier easier, suggesting that selection may favour multigene families being in clustered arrays. It follows from the first conclusion that genes for which mutations may often be dominant or semi-dominant should undergo concerted evolution more commonly than others. By analysis of the mouse knockout database, we show that mutations affecting growth-related genes are more commonly associated with dominant lethality than expected by chance. We predict then that selection will favour homogenization of such genes, and possibly others that are significantly dosage dependent, more often than it favours homogenization in other genes. The first condition is almost the opposite of that required for the maintenance of sexual reproduction according to the mutation-deterministic theory. The analysis here therefore suggests that sexual organisms can simultaneously minimize both the effects of deleterious, strongly synergistically, interacting mutations and those that interact either weakly synergistically, multiplicatively, or antagonistically, assuming the latter class belong to a multicopy gene family. Recombination and an absence of homogenization are efficient in purging deleterious mutations in the former class, homogenization and an absence of recombination are efficient at minimizing the costs imposed by the latter classes.

Keywords: evolution of sex, evolution of ploidy, gene conversion, epistatic interactions, multigene arrays

1. INTRODUCTION

Natural selection has been evoked to explain many features of the structure of genes and genomes; what, for want of a term, might be called 'evolutionary genomics'. For example, selectionist explanations have been provided to account for why some genes might be (i) in multiple copies and others in just a single copy (e.g. Haig 1993; Hurst 1992, 1996), (ii) on one chromosome rather than another (e.g. Hurst 1994; Rice 1992), (iii) on particular parts of a chromosome (e.g. Charlesworth & Hartl 1978), and (iv) why they should be linked to certain other genes (e.g. Bodmer & Parsons 1962; Hutson & Law 1993; Korol *et al.* 1994; Lyttle 1991). Here we ask whether, and under what conditions, the ability to undergo concerted evolution might be an adaptive feature of genomes.

Many multigene families undergo various processes, such as gene conversion and unequal crossing over, that convert copies of the gene into identical copies. Such homogenized families can go to fixation within a species and cause divergence between species. As a consequence, all the copies within the species appear to be 'coordinated' in their evolution and hence the families are said to undergo concerted evolution (reviewed in Elder & Turner 1995). Why should this happen?

*Author for correspondence (l.d.hurst@bath.ac.uk).

The population genetics of homogenization has attracted considerable attention (reviewed in Weir *et al.* 1985). For the most part these analyses examine the population genetic consequences of homogenization. For example, Ohta (1989) considered the changes to the populational load that occur as a consequence of homogenization. She concludes that homogenization always reduces the load. This suggests that competition between reproductively isolated populations, some with homogenization and some without, might typically be won by the populations with concerted evolution. Here we ask a different question, namely whether homogenization might be adaptive at the individual level. More precisely, we ask whether, when a modifier allowing members of a gene family to be homogenized appears in a population, it will spread because of its effect on the gene family.

2. WHY HOMOGENIZATION MAY BE ADAPTIVE

Imagine, for simplicity, a haploid organism with a gene family comprising two genes. Deleterious mutations occur in the genes at some very low rate. Imagine then that one copy has mutated but the other has not. We can now ask what the organism should do. We shall assume that the organism cannot tell which is the mutated copy and which is the wild-type copy. If it could tell, then replacing the mutated with the unmutated copy will always be

Received 25 September 1997 *Accepted* 13 October 1997

advantageous, assuming this to be an error-free process (cf. Bengtsson 1990).

The organism has a choice: either to do nothing, or to randomly pick one of the two copies and replace the other copy with the chosen copy. At an extreme, if two copies of the mutated copy do you no more harm than one mutated copy, then one can only ever gain (in terms of immediate fitness effects) by randomly selecting one of the two copies and replacing the other with this copy. Half the time fitness goes up, half the time fitness stays the same. At the other extreme, if one copy of the mutant does you no harm, but two copies are very deleterious, then it is better, in the short-term, not to risk the production of an array with two mutant copies. Whether homogenization has the effect of immediately increasing or decreasing fitness is therefore dependent on the nature of the fitness interaction between the two copies of the gene.

More generally, we may consider that the fitness of a haploid individual with two mutant versions of the gene is $(1-s)^2+\epsilon$, when $1-s$ is the fitness of the individual with one mutated copy and one wild-type copy (cf. Otto & Feldman 1997). When $\epsilon=0$, interactions are independent, i.e. fitness is multiplicative. For $\epsilon>0$, there is positive epistasis (alias antagonistic epistasis). If $\epsilon<0$, there is negative or synergistic epistasis. If $\epsilon>-s^2$, then we may say that there is weak synergistic epistasis. For the simple two-locus haploid case, it is trivial to show that conversion is immediately advantageous to the individual so long as $\epsilon>-s^2$. Random homogenization of the *Aa* gene pair will result in an individual with a fitness of 1, half the time, and a fitness of $(1-s)^2+\epsilon$, half the time. The average fitness then will be $1-s+(s^2+\epsilon)/2$.

This fitness is higher than that associated with not doing anything (fitness $1-s$) if $\epsilon>-s^2$, i.e. interactions must be weakly synergistic, multiplicative or antagonistic.

A fuller consideration requires analysis of a modifier locus controlling the tendency to convert one gene into another. It is this analysis that we provide below. The modifier could in principle be almost anything: a gene necessary for gene conversion, one affecting recombination rates, if recombination is necessary for homogenization, or simply a structural feature of the DNA of the genes concerned.

3. THE MODEL

Consider a sexual haploid organism (perhaps like *Chlamydomonas*). At an array with two copies of a gene in it, mutation and conversion can take place. We shall start by assuming that recombination does not occur between the two genes in the array. Three states exist for the array: wild-type (*AA*), one copy is mutant (*Aa*), both copies are mutant (*aa*). At a second locus, two alleles exist. *M* is the modifier that allows the organism to select one of the two copies and replace the other copy with the selected copy. The wild-type allele *m* is effectively a null allele. Note that the choice of which allele to select need not be random. So long as bias in selection of which copy to use does not covary with the probability of mutation, such a bias has no effect on the process.

(a) *Step 1. Sex and recombination*

Conversion we shall assume happens just prior to zygote formation. We assume for simplicity that in the presence of the modifier conversion always occurs. Whilst this is an unrealistic assumption, it should not greatly affect the qualitative conclusions. Gametes are then of five types: *MAA* at frequency x_1, *Maa* at frequency x_2, *mAA* at frequency x_3, *mAa* at frequency x_4 and *maa* at frequency x_5. Zygotes form by random mating. After zygote formation, meiosis occurs, and recombination can occur between the modifier and the array at a rate r. Recombination can generate a new form of haploid individual, i.e. one that is *MAa*. From the frequencies x_i, it is possible to calculate the frequencies of these haploid types. These we can label x_{ia}, where x_{6a} is the frequency of *MAa*. These frequencies are:

$$x_{1a}=x_1(x_1+x_2+x_3+(1-r)(x_4+x_5))+rx_2x_3$$
$$x_{2a}=x_2(x_1+x_2+x_5+(1-r)(x_3+x_4))+rx_1x_5$$
$$x_{6a}=rx_4(x_1+x_2)$$
$$x_{3a}=x_3(x_1+x_3+x_4+x_5+(1-r)x_2)+rx_1(x_4+x_5)$$
$$x_{4a}=x_4(x_3+x_4+x_5+(1-r)(x_1+x_2))$$
$$x_{5a}=x_5(x_2+x_3+x_4+x_5+(1-r)x_1)+rx_2(x_3+x_4).$$

(b) *Step 2. Selection between haploid types*

Selection then affects the viability of the six haploid types. We assume that all mutations are deleterious. There are two ways to represent viabilities. Either, as above, to suppose that those that are *AA* are of fitness 1, those that are *Aa* are of fitness $1-s$, and those that are *aa* have fitness $(1-s)^2+\epsilon$. Alternatively, we could consider viabilities 1, $1-Hs$ and $1-s$, H here being a parameter comparable to the penetrance, h, of a mutation at one locus in a diploid model. Given that we wish to compare the results from this analysis with data on dominance in real genes, the latter formulation is preferable. The modifier is associated with a cost, φ. After normalization, to take account of mean populational fitness, the following new frequencies, x_{ib} can be found:

$$x_{1b}=\frac{x_{1a}(1-\varphi)}{\overline{W}},\quad x_{2b}=\frac{x_{2a}(1-\varphi)(1-s)}{\overline{W}},$$
$$x_{6b}=\frac{x_{6a}(1-\varphi)(1-Hs)}{\overline{W}},\quad x_{3b}=\frac{x_{3a}}{\overline{W}},$$
$$x_{4b}=\frac{x_{4a}(1-Hs)}{\overline{W}},\quad x_{5b}=\frac{x_{5a}(1-s)}{\overline{W}},$$

where $\overline{W}$, the mean fitness, is the sum of the numerators.

(c) *Step 3. Mutation*

Mutation can occur at a rate μ. We shall assume that the mutation rate is so low that the probability that two copies of the gene can mutate can be ignored. This being so, the frequency of *AA* gametes is reduced by a factor $(1-2\mu)$, the frequency of *Aa* gametes is reduced by a factor $(1-\mu)$. These latter gametes also increase through mutation of *AA* and *aa* gametes increase in frequency due to mutation of *Aa*. The new haploid frequencies (x_{ic}) can be expressed as functions of x_{ib}:

$$x_{1c}=x_{1b}(1-2\mu),\quad x_{2c}=x_{2b}+\mu x_{6b},$$
$$x_{6c}=x_{6b}(1-\mu)+2\mu x_{1b},$$
$$x_{3c}=x_{3b}(1-2\mu),\quad x_{4c}=x_{4b}(1-\mu)+2\mu x_{3b},$$
$$x_{5c}=x_{5b}+\mu x_{4b}.$$

(d) *Step 4. Conversion*

Finally, if the gametes bear the M allele, conversion, by whatever mechanism, occurs. As noted above, we assume that if the modifier exists conversion will always occur. We can then provide an expression for the frequency of each gamete type in the next generation (x_i') as a function of x_{ic}:

$$x_1' = x_{1c} + \frac{x_{6c}}{2}, \quad x_2' = x_{2c} + \frac{x_{6c}}{2},$$
$$x_3' = x_{3c}, \quad x_4' = x_{4c}, \quad x_5' = x_{5c}.$$

We therefore have an expression for the change in haploid type frequency as a function of its initial value, the mutation rate, the action of selection, and the presence or absence of conversion.

4. MUTATION–SELECTION EQUILIBRIUM

In the absence of the modifier ($x_1=x_2=0$), the three array types will move to a mutation–selection equilibrium. Simultaneously, solving $x_3'=x_3$, $x_4'=x_4$ and $x_5'=x_5$, under the condition that $x_3+x_4+x_5=1$, we can obtain the equilibrium frequencies, these being:

$$x_3^* = \frac{(s-2\mu)(\mu - Hs(1-\mu))}{s((H(2+s)-1)-Hs)},$$
$$x_4^* = \frac{2\mu(2\mu - s)}{s(H((2+s)-1)-Hs)},$$
$$x_5^* = \frac{2\mu^2(Hs-1)}{s((H(2+s)-1)-Hs)}.$$

A mutation–selection equilibrium exists so long as $\mu>0$ and $s>2\mu$. The mean population fitness is $1-2\mu$.

5. THE BEHAVIOUR OF THE MODIFIER

(a) *The cost-free modifier*

We shall start by considering the case of the neutral modifier ($\varphi=0$). The conditions for the invasion of the modifier can be obtained from determining the solution of $\lambda>1$, where λ is the leading eigenvalue of the linearized recursion equation for x_1' and x_2'. For our case,

$$\lambda = a + d - ad + bc$$

where

$$a = \frac{(1-\mu)(x_3^* + x_4^*\left(1 - \frac{r(1+Hs)}{2}\right) + x_5^*(1-r))}{\overline{W}}$$

$$b = \frac{r(1-\mu)(x_3^* + x_4^*\left(\frac{1-Hs}{2}\right))}{\overline{W}}$$

$$c = \frac{\mu x_3^* + x_4^*\left(\mu + \frac{r(1-m-Hs(1+\mu)}{2}\right) + x_5^*(\mu(1-r) + r(1-s))}{\overline{W}}$$

$$d = \frac{x_3^*(1 - r(1-\mu) - s(1-r)) + x_4^*\left(1 - s(1-r) - \frac{r(1-\mu+Hs(1+\mu))}{2}\right) + x_5^*(1-s)}{\overline{W}}$$

and where $\overline{W}$ is the mean population fitness at mutation–selection equilibrium. The condition resolves to:

$$r < \frac{\mu - H(2\mu - s + \mu s)}{(1-\mu)(1-H(2-s))}.$$

A graphical representation of this is given in figure 1. Two special cases are noteworthy. If the modifier is in perfect linkage with the array being modified ($r=0$), then we obtain the condition that $s>2\mu$ must hold. This, note, is the condition for the existence of mutation–selection equilibrium. We therefore conclude that if a mutation–selection equilibrium exists, the modifier will always spread if it is perfectly linked. Considering the other extreme of free recombination, where $r=0.5$, we obtain the condition

$$H > \frac{1-3\mu}{2-\mu(6+s)+s}.$$

For small values of μ, this approximates to $1/(2+s)$. For the case where s is small, this is equivalent to the condition, derived above (§§ 2), that $H>1/2$ (i.e. $\epsilon > -s^2$) must hold. This confirms the verbal argument that homogenization is beneficial in the absence of strong synergistic epistasis between the genes. As s increases, the necessary minimum value of H approaches 1/3.

These results may be understood by considering that the modifier spreads due to linkage disequilibrium between itself (M) and the advantageous AA pair. Conversion initially forces disequilibrium. Selection against the deleterious allele strengthens the disequilbrium, while recombination weakens it. Hence, low rates of recombination and high values of s, both favour homogenization. The verbal argument presented in §2 misses the alteration of linkage disequilibrium due to the action of selection and recombination. Hence, the simple fitness argument is only in agreement with the modifier analysis when $r = 0.5$ and s is small.

(b) *The costly modifier*

When invasion is possible, simulation reveals that it is fastest when linkage is tight. As might then be expected, if modification is costly ($\varphi>0$), tight linkage will be favoured (figure 2). Indeed, if modification is costly, even if linkage is perfect, the condition for the invasion of the modifier resolves simply to

$$\varphi < \frac{\mu}{1-\mu}.$$

So the maximum cost is of the order of the mutation rate and is independent of both H and s. For higher recombination rates, the maximum cost is lower still (figure 2). Tight linkage between the modifier and the array will hence be selectively favourable. Alternatively, we might assume that the maximum cost that the modifier can suffer is proportional to the strength of selection favouring an otherwise cost-free modifier. This being so, we might conclude that if selection on the modifier is weak (when μ is low and s is low) and drift can thus oppose the invasion of the modifier, tight linkage between the modifier and the array will be necessary for the spread of the modifier.

7. LINKAGE BETWEEN THE VIABILITY LOCI

So far we have assumed that the two viability loci (V_1 and V_2) are perfectly linked in an array. What if

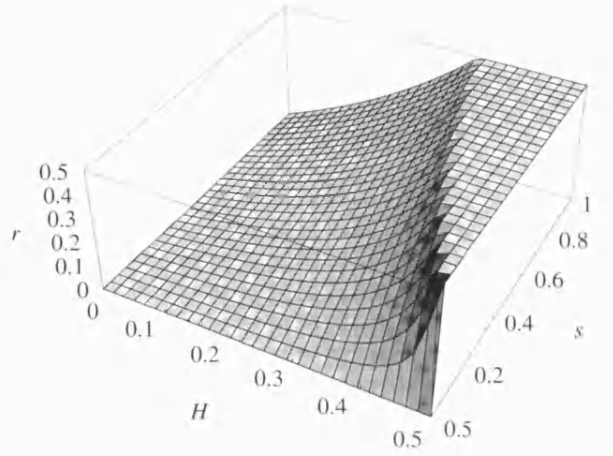


Figure 1. The maximum value of the recombination rate between the modifier and the array, consistent with the invasion of the neutral modifier causing conversion, as a function of effective penetrance (H) of the mutant version of the allele and the selective coefficient (s). This is for the example $s=0.1$ and for $\mu=0.0001$. Note that as H decreases, so the necessity for tight linkage becomes all the more important. Likewise, as s decreases, so does the necessity for linkage increase. These results were confirmed by simulation. The form of the curves is not very sensitive to variation in μ, so long as a mutation–selection equilibrium exists.

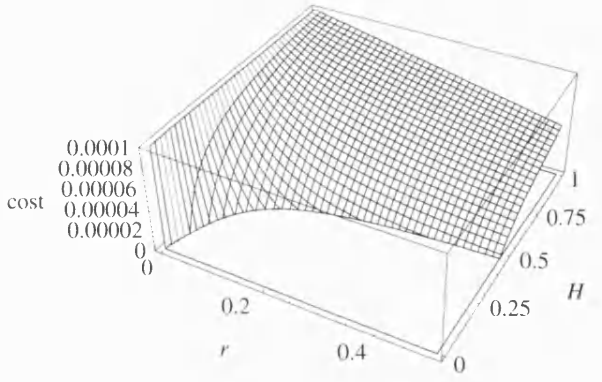


Figure 2. The maximum cost of modification consistent with invasion of the modifier, as a function of the recombination rate (r), and penetrance (H). For this example $\mu=0.0001$ and $s=0.1$.

recombination could occur between these loci as well? Will this make the invasion conditions of the modifier more or less stringent? To investigate this problem we have extended the recursions to allow for recombination between the two viability loci (see Appendix 1).

Our simulations produce two notable conclusions (figure 3). If recombination between the modifier and the first viability locus (V_1) is free ($r_1=0.5$), then recombination (V_2) between the two viability loci makes no difference to the results. The analytical solutions given above for $r=0.5$ are therefore applicable. This is intuitively reasonable. If there is free recombination between M and V_1, then recombination between V_1 and V_2 cannot make the system even more randomized. However, if recombination between the modifier and the first viability locus is not free (i.e. $r_1<0.5$), then lower values of r_2 ease the invasion of the modifier. Invasion is easiest when $r_1 = r_2 = 0$, i.e. the case of perfect linkage given above. We conclude that if homogenization is favoured, selection will also favour linkage between the genes.

8. DISCUSSION

(a) *Conclusions and predictions*

It has been shown that a cost-free modifier controlling the ability of an array to homogenize can spread under particular conditions. If one mutation has little effect, but the second mutation drastically reduces fitness, then a subtle conflict of interests occurs: closely linked modifiers favour homogenization while more distant modifiers do not. If, however, one mutation has a large effect, proportional to that of the second, then a modifier can invade regardless of linkage. Invasion is most likely if the mutations are of large effect. If the modifier is costly, then tight linkage and antagonism between mutations both increase the probability of invasion.

These results can be contrasted with the analysis of the effects of homogenization on load (Ohta 1989) which find that homogenization always reduces the load. In the long-term, then, homogenization is uniformly beneficial to a lineage. Note that the analysis of the perfectly linked modifier is essentially the same as the load argument, and hence both agree that homogenization is beneficial. The modifier analysis shows, however, that homogenization is beneficial in the short-term only under special conditions, i.e. only if interactions between deleterious mutations are not strongly synergistic can the modifier take any linkage arrangement with the array.

We have not considered the effects of homogenization on advantageous alleles. The condition that deleterious mutations must not be strongly synergistic, could alternatively be stated, as regards advantageous mutations, that homogenization may be adaptive if most advantageous mutations are recessive (see, also, Slatkin 1986). We may therefore expect that homogenization will promote the spread of advantageous recessives (for a possible example, see Inomata & Yamazaki 1996). Given, however, that the analysis for a recessive advantageous gene will depend greatly on ploidy, we shall not discuss the problem further.

From the above conclusions we can attempt to formulate predictions about which organisms, and which genes in those organisms, are most likely to undergo homogenization. Whether the load arguments make any qualitatively different predictions to the modifier argument depends critically on what the modifiers are and how they are linked to the genes being modified.

(i) *Prediction 1. Homogenization should be most frequent in genes for which mutations are dominant*

That a freely recombining modifier of homogenization is favoured by selection when interactions between mutations are not strongly synergistic may provide the strongest prediction of the model. This condition, that $H<1/2$, is equivalent to the case, in the normal diploid condition, for a mutation to be partially or wholly dominant. However, highly deleterious mutations tend to be recessive and slightly deleterious ones tend to be partially

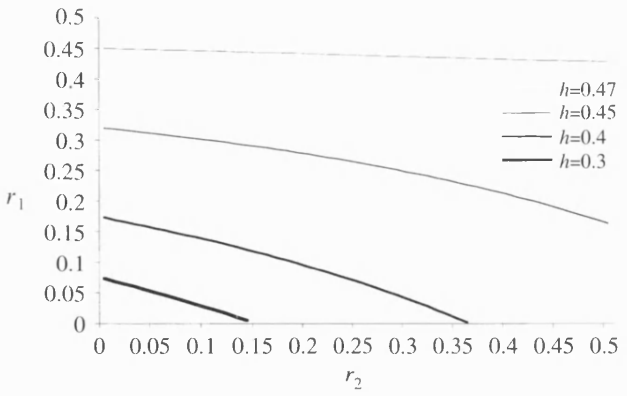


Figure 3. Critical minimum values for the rate of recombination between the modifier and the first viability locus (r_1) and between the two viability loci (r_2) for four values of the penetrance (H). These plots are for the condition where $\mu = 0.0001$ and $s=0.1$. These results were obtained by simulation. The simulations were allowed to run to a mutation–selection equilibrium prior to the introduction of a rare modifier. The modifier is introduced in linkage equilibrium.

recessive (Kacser & Burns 1981). The necessary conditions seem not then to apply to most deleterious mutations. From this we conclude either (1) that homogenization is not adaptive in the short-term, or (2) that the modifiers are usually linked, or (3) that multicopy genes that become homogenized are an unusual class of genes in which mutations may often be at least partially dominant.

It is difficult at present to arbitrate between the three possibilities. However, we wish to suggest that the third possibility may be correct. While most mutations are recessive, mutagenesis studies in many organisms indicate that possibly as many as 10% are dominant or semidominant (Wilkie 1994). Haploinsufficiency is implicated in at least 35 human disease syndromes (Fisher & Scambler 1994) and at least 2000 human genetic disorders are recognized for which the phenotypic penetrance is non-zero (see Online Mendelian Inheritance in Man: `http://www3.ncbi.nlm.nih.gov/omim`). Although referred to as dominant disorders, h is rarely, if ever, equal to unity. Often symptoms are apparent in heterozygotes, but not the full-blown symptoms as witnessed in the homozygotes. Note, however, that for the model the important value is the penetrance with respect to fitness. Heterozygotes may not have the full-blown disease, but in conditions of strong competition they may be liable to die or not reproduce. Their fitness may hence tend towards zero and h (as regards fitness) will thus approach unity, while h, as regards phenotype, may be comparably low.

What sort of genes are likely to have high values of h with respect to fitness, and are multicopy genes more prone to such mutations? Analysing values of h pertinent to fitness rather than phenotype is difficult. We shall instead ask whether there are any generalizations about genes for which mutations are phenotypically dominant (or partially so)? We have searched the mouse knockout database (`http://biomednet.com/cgi-bin/mko/mkobrwse.pl`) to find those genes for which the knockout is lethal in the heterozygous condition (i.e. the stringent case that $s=1$ and $h=1$). We found only eight (the database describes around 250 single-gene knockouts resulting in inviability and over 500 that are viable). These are: (1) core-binding factor alpha 2; (2) glutamate receptor, ionotropic, AMPA2 (alpha 2); (3) hepatocyte growth factor; (4) neural cell adhesion molecule, (5) orthodenticle homologue 2; (6) transforming growth factor, beta-1; (7) vascular endothelial growth factor; and (8) mammalian achaete scute homologue 2 (*Mash2*). There are several more that have major phenotypic effects in heterozygotes, but which are not lethal.

All except one (core-binding factor) of the above eight genes affect growth in one form or another. This might not appear remarkable until one realizes that only about 4% of murine knockouts affect growth (Barlow 1995). There is therefore a vast excess of growth-associated genes in the dominant lethal knockout class. There may be a trivial explanation for this finding, namely the genes may be imprinted (hence half the time the knockouts will be hemizygotes for the knockout). One might suspect this given that there are more imprinted genes affecting growth than one would expect by chance (Barlow 1995). However, *Mash2* is the only one in the list known to be imprinted (Guillemot *et al.* 1995). Knockouts of this are only inviable if maternally inherited. We know of no similar evidence to implicate any of the other genes in imprinting. Importantly, too, some genes known not to be imprinted do have major growth effects in the heterozygous knockout condition (e.g. *Igf1*; Liu *et al.* 1993). Of the 35 human disorders associated with haploinsufficiency (Fisher & Scambler 1994), none are known to be associated with imprinting.

If the over-abundance of growth-associated genes in the set of dominant lethal knockouts is not due to their being imprinted, why else might this result be found? It seems reasonable to suppose that dosage sensitivity might be what one would expect *a priori* of genes with effects in the heterozygous condition (Wilkie 1994). There is considerable evidence that many growth factors are dosage dependent. Incremental over-expression of insulin-like growth factor II, shows a large and positively correlated response of embryonic size to dosage (Sun *et al.* 1997). Similarly, over-expression of growth factors (and their suppressors) frequently results in both alterations of dimensions and lethality (Hurst & McVean 1997). We might then predict that dosage-sensitive genes, of which growth factors would be one important class, should more commonly undergo concerted evolution than others (for discussion of which genes might be dose sensitive, see Fisher & Scambler 1994). One can make an anecdotal case that this might be true. Ribosomal DNA, splice factors, and ubiquitination genes all have been reported as undergoing homogenization (e.g. Hillis *et al.* 1991; Liao *et al.* 1997; Pavelitz *et al.* 1995; Schlotterer & Tautz 1994; Zhou & Ragan 1995) and are all produced at very high levels.

Assuming that dosage-sensitive genes are likely to benefit from homogenization, we can also predict that concerted evolution will probably be an extremely common phenomenon amongst repeat genes that share the same function. This comes from the understanding that selection will favour the duplication of a gene, if the effect of the gene is dosage dependent. Hence, we may expect, *a priori*, that multicopy genes are more likely to be those that are dose sensitive, and hence could benefit from homogenization.

The above proposition is testable. Not all gene families need be dosage dependent. Some multicopy families do not have all members expressed in the same tissue. Instead, the numerous copies each have tissue-specific activity. We may then predict that gene families in which all (or most) copies are expressed synchronously in the same cell (e.g. rDNA) should have dominant mutations more often than those in which different members of the gene family have tissue-specific expression, and hence do not coexpress in the same cell.

(ii) *Prediction 2. Homogenization should be more common when recombination rates are low*

That invasion is always easiest when the array and the modifier are linked suggests that homogenization may be more common for arrays with low recombination rates and that selection may favour multigene families being in linked arrays. It also suggests that organisms with low recombination rates should have higher rates of homogenization. However, these may not be strong predictions, for two reasons. First, recombination may be necessary for homogenization to occur (unequal crossing over, for example, requires crossing over to occur). Second, advantageous mutations have a hard time spreading in asexual populations or in regions of low recombination (reviewed in Hurst & Peck 1996). A modifier linked to an array in a region of low recombination might then be invisible to selection. These caveats aside, it is noteworthy that if one array has the ability to undergo homogenization, selection could favour the linkage of that array with others that would benefit from homogenization. Suggestively, this clumping of different arrays is a commonly evolved trait (Drouin & de Sa 1995).

(iii) *Prediction 3. Homogenization should be most common in organisms with large population sizes*

The maximum cost that the modifier can suffer and still invade is of the order of the mutation rate of each gene in the array. This suggests that even if interactions are strongly antagonistic (dominant), if selection is weak, homogenization may not be favoured. As the efficiency of selection is dependent upon the population size, we may conclude that homogenization is more likely to be adaptive in species whose population size is large. Rough calculations (not shown) suggest that it is unlikely that mammalian populations are large enough to allow selection to favour a modifier affecting only one gene array. If, however, one modifier can affect many arrays simultaneously, then its spread may well be adaptive. Under this circumstance, however, it is unlikely that the modifier could be simultaneously linked to all arrays (but see Drouin & de Sa 1995). Hence, there may be differences between taxa in the linkage of genes undergoing homogenization, with those in small populations being unlinked more commonly.

(b) *Homogenization, sex and ploidy*

As might perhaps be expected, our results are similar to those for the non-invasion of a modifier of diploidy entering a haploid population (Otto & Goldstein 1992). Diploidy may be advantageous because deleterious mutations are masked. Conversely, haploidy and homogenization are favoured if masking is not adequately effective. Another way to read this conclusion is to suppose that a diploid organism can treat different genes differently. Those for which masking is beneficial can be left unhomogenized. Those for which masking is deleterious can be homogenized.

The theory outlined here is important in the context of theories of sex. Kondrashov (1988) has shown that if interactions between deleterious mutations are synergistic, then sex will be favoured, so long as the mutation rate is high enough, as it increases variance between progeny in the degree of genome contamination (see also Barton 1995; Charlesworth 1990; Kondrashov 1984). What, one might then ask, should an organism do if some interactions are synergistic but others are not (or only weakly so)? More generally, what will happen if there is variance in the form of the ϵ, the parameter defining the form of the epistatic interaction? It has been shown that such variance reduces the parameter space in which recombination is beneficial (Otto & Feldman 1997). For evidence of such variance see Elena & Lenski (1997).

The analysis provided here goes some way to countering this objection to the mutational deterministic theory. Assuming that the genes interacting with $\epsilon > -s^2$ belong to the same gene family, by homogenizing these alone and preventing recombination between the viability loci, a sexual organism can minimize the deleterious effects of any mutations. In contrast, by promoting recombination and preventing homogenization, a sexual organism can minimize the effects of mutations that interact synergistically. It is as though part of the genome was sexual (not homogenized and recombining), and part was asexual (homogenized and not recombining). It seems then that with respect both to haploidy/diploidy and sexual/asexual, the organism can, to coin a term, 'have its cake and eat it'.

We wish to thank Alexey Kondrashov, Eörs Szathmáry, and Gil McVean for comments on an earlier version of the manuscript. L.D.H. is funded by the Royal Society.

APPENDIX 1

Recursions for the three locus model. We consider a gene order $M\ V_1\ V_2$ where V_1 and V_2 are the two viability loci. Recombination occurs between M and V_1 at a rate r_1, and between V_1 and V_2 at a rate r_2.

To avoid confusion with the original model we shall adopt a new nomenclature incorporating the haploid frequencies as terms in z rather than x. There are eight haploid types prior to zygote formation, these being: *MAA* at frequency z_1, *MAa* at frequency z_{21}, *MaA* at frequency z_{22}, *Maa* at frequency z_3, *mAA* at frequency z_4, *mAa* at frequency z_{51}, *maA* at frequency z_{52}, and *maa* at frequency z_6. As in the above model, prior to zygote formation there are no heterozygotes associated with the modifier ($z_{21} = z_{22} = 0$). After sex and recombination the new genotype frequencies become:

$$\begin{aligned} z_{1a} &= z_1[z_1 + z_1 z_3(1 - r_2) + z_4 + z_{51}((1 - r_1)(1 - r_2) \\ &\quad + r_1 r_2) + z_{52}(1 - r_1) + z_6(1 - r_1)(1 - r_2)] \\ &\quad + z_3 z_4 r_1(1 - r_2), \\ z_{21a} &= z_1[z_3 r_2 + z_{51}(r_1(1 - r_2) + (1 - r_1)r_2) \\ &\quad + z_6(1 - r_1)r_2] + z_3[z_4 r_1 r_2 + z_{51} r_1], \end{aligned}$$

$$z_{22a} = z_1[z_3r_2 + z_{52}r_1 + z_6r_1r_2] + z_3[z_4(1 - r_1)r_2 + z_{52}(r_1(1 - r_2) + (1 - r_1)r_2)],$$

$$z_{3a} = z_1[z_3(1 - r_2) + z_6r_1(1 - r_2)] + z_3[z_3 + z_4(1 - r_1)(1 - r_2) + z_{51}(1 - r_1) + z_{52}((1 - r_1)(1 - r_2) + r_1r_2) + z_6],$$

$$z_{4a} = z_1[z_4 + z_{51}(r_1(1 - r_2) + (1 - r_1)r_2) + z_{52}r_1 + z_6r_1(1 - r_2)] + z_4[z_3(1 - r_1)(1 - r_2) + z_4 + z_{51} + z_{52} + z_6(1 - r_2)] + z_{51}z_{52}r_2,$$

$$z_{51a} = z_1[z_{51}((1 - r_1)(1 - r_2) + r_1r_2) + z_6r_1r_2] + z_3z_4(1 - r_1)r_2 + z_{51}[z_3(1 - r_1) + z_4 + z_6 + z_{51} + z_{52}(1 - r_2) + z_6],$$

$$z_{52a} = z_1[z_{52}(1 - r_1) + z_6(1 - r_1)r_2] + z_4[z_3r_1r_2 + z_6r_2] + z_{52}[z_3((1 - r_1)(1 - r_2) + r_1r_2) + z_4 + z_{51}(1 - r_2) + z_{52} + z_6],$$

$$z_{6a} = z_3[z_4r_1(1 - r_2) + z_{51}r_1 + z_{52}(r_1(1 - r_2) + (1 - r_1)r_2) + z_6] + z_{51}z_{52}r_2 + z_6[z_1(1 - r_1)(1 - r_2) + z_3 + z_4(1 - r_2) + z_{51} + z_{52} + z_6].$$

After this the recursions are trivial adaptations of those given in the text and so not presented. *Aa* types have the same fitness and mutation rates as *aA* types, i.e. gene order does not affect these parameters. For results of the simulations see figure 3.

REFERENCES

Barlow, D. P. 1995 Gametic imprinting in mammals. *Science* **270**, 1610–1613.

Barton, N. H. 1995 A general-model for the evolution of recombination. *Genet. Res.* **65**, 123–144.

Bengtsson, B. O. 1990 The effect of biased conversion on the mutation load. *Genet. Res.* **55**, 183–187.

Bodmer, W. F. & Parsons, P. A. 1962 Linkage and recombination in evolution. *Adv. Genet.* **11**, 1–100.

Charlesworth, B. 1990 Mutation–selection balance and the evolutionary advantage of sex and recombination. *Genet. Res.* **55**, 199–221.

Charlesworth, B. & Hartl, D. L. 1978 Population dynamics of the segregation distorter polymorphism of *Drosophila melanogaster*. *Genetics* **89**, 171–192.

Drouin, G. & de Sa, M. M. 1995 The concerted evolution of 5S ribosomal genes linked to the repeat units of other multigene families. *Molec. Biol. Evol.* **12**, 481–493.

Elder, J. F. & Turner, B. J. 1995 Concerted evolution of repetitive DNA sequences in eukaryotes. *Q. Rev. Biol.* **70**, 297–320.

Elena, S. F. & Lenski, R. E. 1997 Test of synergistic interactions among deleterious mutations in bacteria. *Nature* **390**, 395–398.

Fisher, E. & Scambler, P. 1994 Human haploinsufficiency—one for sorrow, two for joy. *Nat. Genet.* **7**, 5–7.

Guillemot, F., Caspary, T., Tilghman, S. M., Copeland, N. G., Gilbert, D. J., Jenkins, N. A., Anderson, D. J., Joyner, A. L., Rossant, J. & Nagy, A. 1995 Genomic imprinting of *Mash2*, a mouse gene required for trophoblast development. *Nat. Genet.* **9**, 235–242.

Haig, D. 1993 Genetic conflicts in human pregnancy. *Q. Rev. Biol.* **68**, 495–532.

Hillis, D. M., Moritz, C., Porter, C. A. & Baker, R. J. 1991 Evidence for biased gene conversion in concerted evolution of ribosomal DNA. *Science* **251**, 308–310.

Hurst, L. D. 1992 Is *Stellate* a relict meiotic driver? *Genetics* **130**, 229–230.

Hurst, L. D. 1994 Embryonic growth and the evolution of the mammalian Y chromosome. I. The Y as an attractor for selfish growth factors. *Heredity* **73**, 223–232.

Hurst, L. D. 1996 Further evidence consistent with *Stellate*'s involvement in meiotic drive. *Genetics* **142**, 641–643.

Hurst, L. D. & McVean, G. T. 1997 Growth effects of uniparental disomies and the conflict theory of genomic imprinting. *Trends Genet.* **13**, 436–443.

Hurst, L. D. & Peck, J. R. 1996 Recent advances in understanding the evolution and maintenance of sex. *Trends Ecol. Evol.* **11**, 46–52.

Hutson, V. & Law, R. 1993 Four steps to two sexes. *Proc. R. Soc. Lond.* B **253**, 43–51.

Inomata, N. & Yamazaki, T. 1996 Adaptive evolution at the molecular level of the duplicated AMY gene system in *Drosophila*. *J. Genet.* **75**, 125–137.

Kacser, H. & Burns, J. A. 1981 The molecular basis of dominance. *Genetics* **97**, 639–666.

Kondrashov, A. 1988 Deleterious mutations and the evolution of sexual reproduction. *Nature* **336**, 435–440.

Kondrashov, A. S. 1984 Deleterious mutations as an evolutionary factor. I. The advantage of recombination. *Genet. Res.* **44**, 199–217.

Korol, A. B., Preigel, I. A. & Preigel, S. I. 1994 *Recombination variability and evolution*. London: Chapman & Hall.

Liao, D. Q., Pavelitz, T., Kidd, J. R., Kidd, K. K. & Weiner, A. M. 1997 Concerted evolution of the tandemly repeated genes encoding human U2 snRNA (the RNU2 locus) involves rapid intrachromosomal homogenization and rare interchromosomal gene conversion. *EMBO J.* **16**, 588–598.

Liu, J. P., Baker, J., Perkins, A. S., Robertson, E. J. & Efstratiadis, A. 1993 Mice carrying null mutations of the genes encoding insulin-like growth factor-I (*Igf-1*) and type-1 Igf receptor (*Igf1r*). *Cell* **75**, 59–72.

Lyttle, T. W. 1991 Segregation distorters. *A. Rev. Genet.* **25**, 511–557.

Ohta, T. 1989 The mutational load of a multigene family with uniform members. *Genet. Res.* **53**, 141–145.

Otto, S. P. & Feldman, M. W. 1997 Deleterious mutations, variable epistatic interactions, and the evolution of recombination. *Theor. Popul. Biol.* **51**, 134–147.

Otto, S. P. & Goldstein, D. B. 1992 Recombination and the evolution of diploidy. *Genetics* **131**, 745–751.

Pavelitz, T., Rusche, L., Matera, A. G., Scharf, J. M. & Weiner, A. M. 1995 Concerted evolution of the tandem array encoding primate U2 snRNA occurs *in situ*, without changing the cytological context of the RNU2 locus. *EMBO J.* **14**, 169–177.

Rice, W. R. 1992 Sexually antagonistic genes—experimental evidence. *Science* **256**, 1436–1439.

Schlotterer, C. & Tautz, D. 1994 Chromosomal homogeneity of *Drosophila* ribosomal DNA arrays suggests intrachromosomal exchanges drive concerted evolution. *Curr. Biol.* **4**, 777–783.

Slatkin, M. 1986 Interchromosomal biased gene conversion, mutation and selection in a multigene family. *Genetics* **112**, 681–698.

Sun, F. L., Dean, W., Kelsey, G., Allen, N. & Reik, W. 1997 Transactivation of *IGF2* results in Beckwith Wiedemann syndrome phenotypes. *Nature* **389**, 809–815.

Weir, B. S., Ohta, T. & Tachida, H. 1985 Gene conversion models. *J. Theor. Biol.* **116**, 1–8.

Wilkie, A. O. M. 1994 The molecular basis of genetic dominance. *J. Med. Genet.* **31**, 89–98.

Zhou, Y. H. & Ragan, M. A. 1995 Characterization of the polyubiquitin gene in the marine red alga *Gracilaria verrucosa*. *Biochem. Biophys. Acta* **1261**, 215–222.

Research Paper 3. Vertebrate genome evolution: a slow shuffle or a big bang?

Nick Smith, Rob Knight and Laurence Hurst (1999)
Bioessays **21** 697-703.

Vertebrate genome evolution: a slow shuffle or a big bang?

Nick G.C. Smith,[1]* Robert Knight,[2] and Laurence D. Hurst[1]

Summary

In vertebrates it is often found that if one considers a group of genes clustered on a certain chromosome, then the homologues of those genes often form another cluster on a different chromosome. There are four explanations, not necessarily mutually exclusive, to explain how such homologous clusters appeared. Homologous clusters are expected at a low probability even if genes are distributed at random. The duplication of a subset of the genome might create homologous clusters, as would a duplication of the entire genome. Alternatively, it may be adaptive for certain combinations of genes to cluster, although clearly the genes must have duplicated prior to rearrangement into clusters. Molecular phylogenetics provides a means to examine the origins of homologous clusters, although it is difficult to discriminate between the different explanations using current data. However, with more extensive sequencing and mapping of vertebrate genomes, especially those of the early diverging chordates, it should soon become possible to resolve the origins of homologous clusters. *BioEssays* 21:697–703, 1999.

Introduction

It is often found that if one considers a cluster of genes in eukaryotes, then the homologs (assigned on the basis of sequence similarity) of those genes often form another cluster on a different chromosome.[1] The regions encompassing the Hox clusters are a well-known example.[2] We use the term "homologous clusters" to indicate such sets of clusters. At present, it is not known either how or why homologous clusters came to exist. Here we review the evidence with regard to a number of hypotheses, and tentatively suggest directions for future research.

The most popular explanations for the origins of homologous clusters invoke block duplication, either of the entire genome, or of a subset of the genome.[3] For example, it now appears that the Hox clusters have arisen as the result of three separate duplication events.[4] Such explanations suggest that homologous clusters arise in a single event (a "big bang"), and have not been obscured by the subsequent effects of mutation and selection.

An alternative hypothesis suggests that homologous clusters may have arisen because of their adaptive value. A recent article by Hughes[5] on the phylogeny of genes in one family of clusters extends previous results[6] and claims to provide support for the adaptive theory, under which gene duplication and homologous cluster formation need not occur simultaneously. Homologous genes are generated by duplication, and then a series of genome rearrangements take place, which create adaptive clusters of linked genes (a "slow shuffle"). If selection favours similar sets of homologous genes come together, then homologous clusters are the result.

Coincidence may seem a poor explanation of homologous clusters, but until we know how prevalent homologous clusters are, then the effects of chance cannot be discounted. After all, given the huge number of genes in the vertebrate genome, and the existence of many families of homologous genes, the existence of a few homologous clusters is only to be expected.

Alternative models applied to a specific case

We shall consider alternative explanations for the homologous clusters recently investigated by Hughes.[5] The clusters

[1]Department of Biology and Biochemistry, University of Bath, Bath, UK.
[2]Division of Zoology, School of Animal and Microbial Sciences, University of Reading, Reading, UK.
*Correspondence to: Nick Smith, School of Biological Sciences, University of Sussex, Brighton, BN1 9QG, UK. E-mail: n.g.c.smith@sussex.ac.uk

in question are found on chromosomes 6 (6q21.3), 9 (9q33–34), and 1(1p21–25 1p13) in humans, which are homologous with those on mouse chromosome 17, 2, and 1 respectively.[7,8] The regions on human chromosome 6 and mouse chromosome 17 are both near the major histocompatibility complex (MHC). There is some evidence for a fourth cluster in humans on chromosome 19.[7,8] These homologous clusters contain sequenced genes from nine families: Retinoid X Receptor (RXR) genes, Collagen (COL) genes, ATP-binding Cassette Transporter (ABC) genes, Proteasome Component β (PSMB) genes, Notch (NOTCH) genes, Pre-B-cell Leukemia Transcription Factor (PBX) genes, Tenascin (TEN) genes, C3/C4/C5 Complement Component (C3/C4/C5) genes, and Heat Shock Protein 70 (HSP70) genes (gene family abbreviations in parentheses as given by Hughes[5]). Coincidence seems unlikely to explain the existence of four homologous clusters, although we cannot rule out this possibility by considering a single case.

Such a pattern of homologous clusters can be explained by two or three independent duplication events, depending on whether the duplications were of the entire genome or a subset of the genome (see Fig. 1A). An initial duplication event affecting a single chromosomal region (D1) yielded two duplicated regions in different parts of the genome. Duplications of both of these regions then produced the homologous clusters now found on human chromosomes 1, 9, 6, and 19. The second stage of duplications could have been a single polyploidisation event (D2 and D3 simultaneous) with two duplications required overall, or two separate block duplications (D2 and D3 separate) with three duplications required overall.

Hughes reasoned that if block duplication explains the patterns, then a phylogeny of the genes should reflect this.[5] For simplicity we shall call the genes from the nine gene families *A* to *I*, with a suffix being used to indicate the human chromosome (i.e., *A6* is the *A* gene on human chromosome 6). If block duplication explains the pattern then a phylogenetic analysis should reveal the same time to common ancestor for *A6* and *A9* (i.e., time back to initial duplication event), as would be found from a phylogenetic reconstruction of the time of common ancestry of *B6* and *B9*, *C6* and *C9*, and so on. Further if the block duplication were the result of tetraploidization at the base of the vertebrates, as often conjectured,[3] there should be no non-vertebrate genes appearing in the phylogeny within the groupings defined by branches for the genes on chromosomes 6, 19, 9, and 1 (see Fig. 1B). All the invertebrate genes should appear as a mass sister grouping to the vertebrate copies.

If, alternatively, *A6* and *A9* diverged at the origin of the eukaryotes (for example) and independently came into a cluster with *B* genes, *C* genes etc., then human *A6* might well be more closely related to the *Drosophila* homolog of *A6* than to human *A9* (see Fig. 1C). At the same time, *B6* and *B9* could have had a different time to a common ancestor altogether.

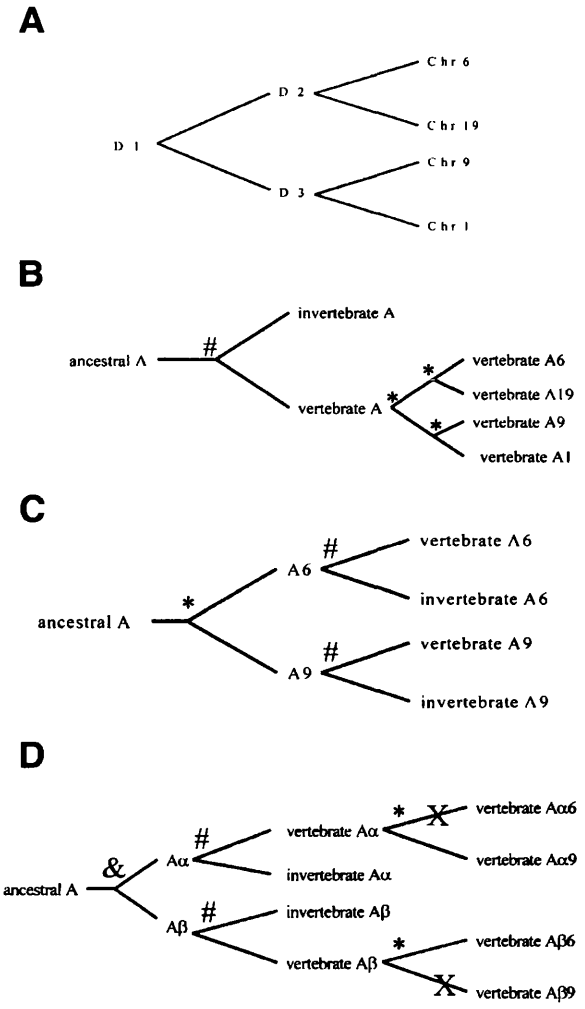


Figure 1. A scheme for the origin of duplicated regions on human chromosomes 6, 19, 1, and 9 is given in **A**. Three duplication events are given by D1, D2, and D3. The ancestral chromosome duplicated to give two chromosomes, both of which then duplicated to give the four present day chromosomes. **B** illustrates the fact that if related gene clusters have arisen in the vertebrate lineage, then the divergence times of such regions in vertebrates should be after the origin of the vertebrates. The split of the vertebrates and invertebrates is represented by #, and block duplications are given by *. **C** shows that if gene duplications occurred before the origin of the vertebrates, then invertebrate genes could cluster with vertebrate genes. **D** demonstrates how gene loss can lead to misleading estimates of divergence time. If the two remaining copies of gene A in the vertebrates are Aα9 and Aβ6, then their time of divergence will be prior to the time of the vertebrate-invertebrate split, despite block duplications involving A occurring in the vertebrate lineage. An ancient tandem duplication is given by &, and X portrays gene loss.

What patterns of molecular phylogeny did Hughes find for the gene families in the homologous clusters? For three to five of the genes, early vertebrate duplication (prior to the divergence of jawed and jawless vertebrates) is found, as expected of block duplication and/or early vertebrate polyploidization. But the remaining four showed wide ranging patterns, with divergences all before the origin of the vertebrates, some going back to the common ancestor of animals and fungi and one even going back as far as the eukaryote-eubacterial common ancestor.

This, Hughes argues, "decisively rejects" the hypothesis of block duplication. However the null hypothesis that he rejects supposes that no gene loss follows a block duplication. Imagine instead that gene *A* in the ancestral vertebrate genome existed in two tandem copies, rather than one. These we can call *A*α and *A*β (see Fig. 1D). The tandem duplication event might have occurred any time prior to the vertebrate expansion: at the origin of eukaryotes, at the origins of animals, whenever. We consider a tandem duplication event because the two duplicates would remain closely linked. Now imagine that after the block duplication *A*α is lost from the pre-chromosome 6, but that on the pre-chromosome 9 *A*β is lost. Under the methodology employed by Hughes such a pattern of events will lead to rejection of the hypothesis of early vertebrate block duplication. This is because when we do a phylogeny of the *A* genes, the one left on chromosome 6 (*A*β) has a most recent common ancestor with that on chromosome 9 (*A*α) sometime well before the vertebrates, at the time of the duplication of the original A gene into *A*α and *A*β.

Such pre-vertebrate tandem duplication events followed by reciprocal loss (*A*α on one and *A*β on the other) yield a model consistent with vertebrate block duplication (see Fig. 2A and B).[7,8] The block duplication model also requires that the clusters on 1 and 9 have a recent common ancestry and that the loss events occurred before this duplication. The phylogenetic evidence is consistent with this (note similar patterns of gene loss on chromosomes 1 and 9 in Fig. 2B).

A similar argument invoking gene loss has been used to account for the inconsistent phylogenetic data of related clusters of insulin group genes and aromatic amino acid hydroxylase genes.[9] Duplication dates of some of these genes prior to the origin of the vertebrates would seem, at first, to invalidate the hypothesis of block duplication at the base of the vertebrates. When gene loss and gene conversion are taken into account, however, the results are consistent with block duplication.

Discriminating between alternative models in the specific case

From the analysis above it appears that a model of block duplication that incorporates gene loss cannot be rejected, even if a model without gene loss can be rejected. Which model is the most consistent with existing data? Hughes rightly complains that advocates of block duplication often come up with rather elaborate post-hoc hypotheses to argue their way out of tight corners. It is certainly awkward to have to propose the existence of presently unidentified duplications. But then how parsimonious is it to suppose that three to five of the clustered genes duplicated at the same time, but that there was no block duplication?

The assumption of no gene loss following gene duplication seems unreasonably prohibitive. The generation of identical duplicate genes implies that the loss of one of the two copies is unlikely to have deleterious effects. From the distributions of the number of genes in human gene families Nadeau and Sankoff have estimated that the rate of gene loss is about the same as the rate of functional retention and divergence.[10] In other words about 50% of duplicates are driven to non-functionality by mutation. There are two possible routes for functional duplicates to be retained. New functions may evolve, especially if many genes are multifunctional.[11] Alternatively, the duplicates may be retained as a buffer against developmental error.[12,13] However, these routes are unlikely to be available to all genes, which means that a certain proportion of duplicates will be lost. Further evidence of gene loss comes from Hox cluster genes: different paralogs have been lost in the different clusters, and furthermore patterns of gene loss differ between lineages (e.g., man and pufferfish[14]). In light of the scheme we presented above to suggest how gene loss can lead to the misallocation of paralogy (Fig. 1D), the fact that gene loss does occur highlights the need for caution in assigning paralogy without adequate phylogenetic evidence. By adequate phylogenetic evidence we mean that one should include groups of organisms that enable the reliability of paralogy assignments to be tested (such as early chordates, see below).

In favour of the block duplication hypothesis, the timings of the gene duplication and gene loss events proposed in Figure 2B seem reasonable. If ancient tandem duplications have been maintained for many millions of years, we expect many of the tandem duplicates to have evolved different functions. Only when further duplications occur can gene loss be expected. The reciprocal nature of gene loss, whereby one α copy and one β copy is lost, also seems reasonable given the different functions of the α and β copies.

Additionally, if the criterion in model discrimination is parsimony, it should be noted that Hughes' alternative model seems even more ornately baroque than the block duplication model, and gene loss. One must suppose that five of the genes (say *A*, *B*, *C*, *D*, and *E*) underwent duplication at the same time (in a block or separately) but that the others (*F*, *G*, *H*, and *I*) had duplicated earlier. If the first five were block duplicated then one must suggest that one copy of *F* goes to each of the four homologous clusters. If there was no block duplication one must suggest that all nine gene families

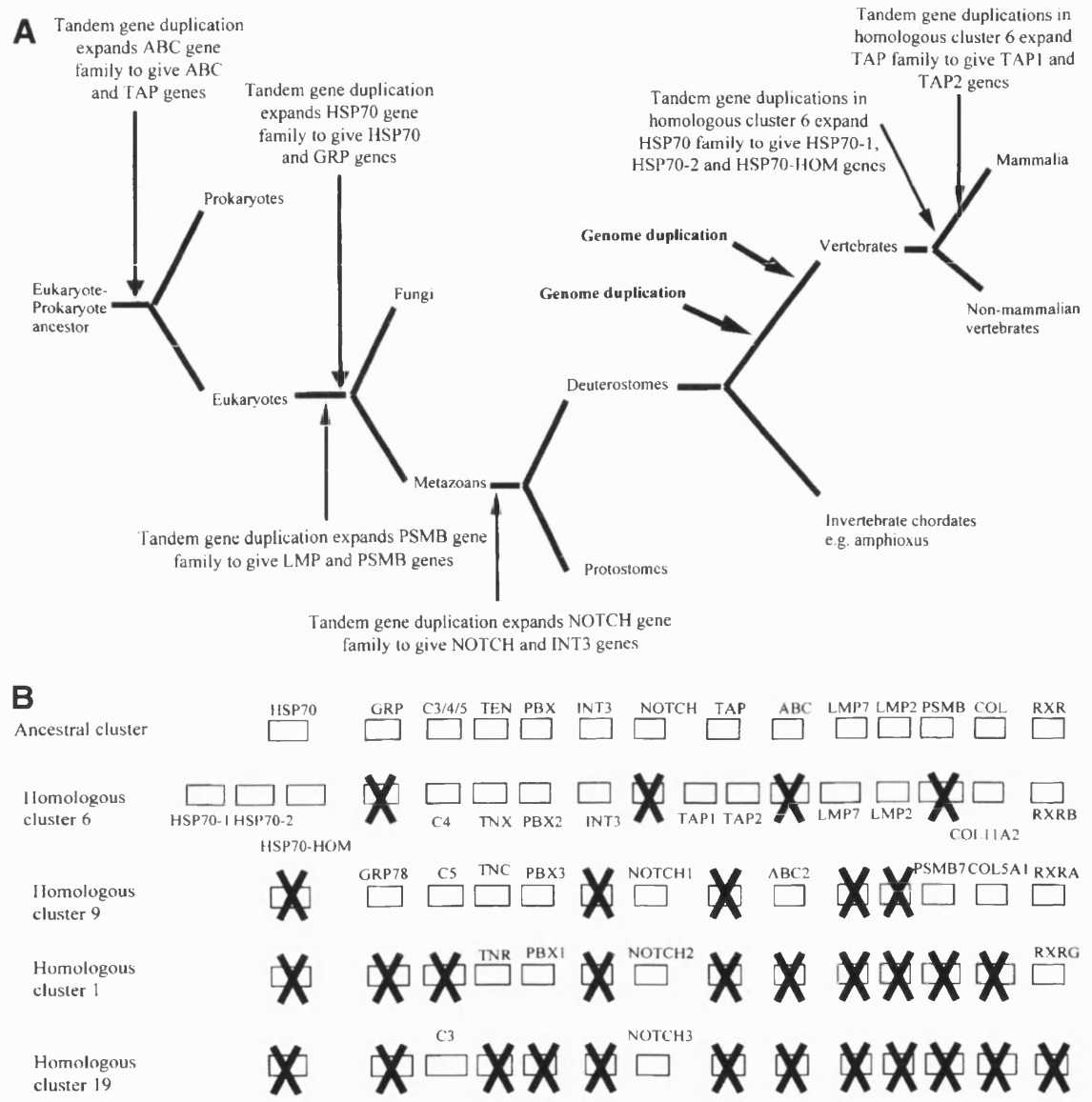


Figure 2. A model for the origin of the homologous clusters seen in humans at 1p13, q21–25, 9q33–34, 6p21.3, and 19p13.3, involving chromosomal duplications (**A**), tandem gene duplications (A), and gene loss (**B**). Early tandem duplications which occur prior to the vertebrates can be estimated from the topology of the gene trees. An initial duplication of the ancestral cluster in the vertebrate ancestor would generate the precursors for homologous clusters 6/19 and 1/9, and a second round of duplications at a later point generated the homologous clusters at 1, 9, 6, and 19 (see A). After each chromosomal duplication, extensive gene loss creates the pattern observed today (see B). Crosses indicate that either the gene has been lost or not yet found (in particular, many genes may be undiscovered within homologous cluster 19). The gene family abbreviations are given in the text.

independently came together with at least one member of each gene family in each homologous cluster.

There is one issue on which block duplication and adaptive models of homologous cluster formation are agreed: a cluster of genes must have formed at some point. The question then becomes whether the cluster came together once (block duplication hypothesis) or several times (adaptive hypothesis). Hybrid models, in which the block duplication and adaptive hypotheses are combined, are also possible. The original cluster might have come together for adaptive reasons, and then block duplications might have created the homologous clusters. Alternatively, block duplica-

tions might have created many homologous clusters, of which only those favoured by selection now remain (L. Lundin, personal communication).

Why might the original cluster have formed, or why might homologous clusters be maintained? It might be selectively advantageous for genes whose products interact to be linked, perhaps in order to enable better regulation. Such an interaction of genes within an homologous cluster appears to apply to the present case. The Proteasome Component β gene family contains the two genes *LMP2* and *LMP7*. These combine to make the LMP+ proteasome that breakdown proteins into peptides that are presented by MHC class I molecules. The peptides are transported across the endoplasmic reticulum (ER) membrane by a dimeric transporter Tap. The *Tap* gene belongs to the ATP-binding Cassette Transporter gene family. Of course it may be that this relationship between members of the homologous clusters may be just a coincidence.

However, the combination of functional interaction and physical linkage is by no means unique (for a review see Reference 15). For example, in *Caenorhabditis*[16] two different enzymes are needed for trimerizing collagen and these are encoded within one operon. The best data for selection on linkage come from meiotic drive genes[17] and supergene complexes.[18] The problem with such examples is that we are aware of no work that reliably predicts the regularity of such coincidences.

As for the evolutionary history of the homologous clusters studied by Hughes, we seem to be left with two models, one of block duplication and one of selection, both of which are really quite complicated. How can we discriminate between the two alternatives? One potentially informative piece of evidence concerns the form of the ancestral vertebrate clusters. Under the block duplication and gene loss model tandem duplications prior to the origin of the vertebrates are required for four gene families (Fig. 2A and B). Tandem duplications lead to close linkage of duplicates, which makes the misallocation of paralogy more likely (Fig. 1D). But the alternative model of ancient duplications followed by gene shuffling makes no specific prediction of tandem duplicates within the gene clusters.

Hence, if one examines the ancestral cluster in an early diverging chordate such as amphioxus or a tunicate and finds tandem duplications within the cluster, then that is good evidence in favour of the block duplication and gene loss model. Furthermore if one assumes that gene conversion has not taken place, then the time of common ancestry of such tandem duplicates and, say, human genes should be the time of the divergence of the early chordates. On the other hand, the absence of tandem duplicates in early chordates might be due to gene loss, and does not necessarily favour the adaptive hypothesis. Tandems duplicates may well have been maintained for many millions of years (from the ancient duplication event until the divergence of the early chordates), and then lost in early diverging chordates due to changing selective pressures, if the biotic environment was rapidly evolving at the time of the vertebrate radiation.

General explanations for the evolution of homologous clusters

In the specific case described earlier, our basis for discriminating between the alternative models of homologous cluster evolution was how well the models fitted the predicted molecular phylogenies. The problem with this general method lies in the assumption that our predicted molecular phylogenies are correct. We have already shown how the correct identification of paralogs is complicated if gene duplication is followed by random gene loss. In addition, tree reconstruction is sensitive to differences in evolutionary rates. Long branch attraction, whereby the longest branches are inferred as outgroups irrespective of the true phylogeny, can be the result of variation in evolutionary rates.[19] Furthermore, the detection of fast evolving sequences is hampered by mutational saturation.[20] Differences in rates of evolution between genes within a homologous cluster could be the result of differences in mutation or selection. Given that one of the ways in which newly formed duplicates survive is through the acquisition of new roles, directional selection is likely to have affected many of the genes in the homologous clusters. Even homeobox genes, commonly thought to be highly conserved and subject to only purifying selection, have been affected by positive selection.[21]

Putting aside problems of tree reconstruction, is there any way we can decide between the different hypotheses for the general phenomenon of homologous clusters? Until we attempt to address general explanations for the evolution of homologous clusters our understanding is restricted to the level of anecdote. In principle, it might be possible to evaluate the likelihoods of evolutionary histories such as that presented in Figure 2B, if one had sufficient information on mutational biases. It should be possible to determine the relative frequencies of gene loss, different sorts of translocations, and different sorts of duplications. For example, comparative mapping data have been used to estimate distributions of rearrangement breakpoints.[22] From such mutational data, it should be possible to predict the null (coincidence) expectation for the frequency of homologous clusters. Only if the null expectation is significantly less than the observed frequency need we consider any further explanations.

However, the mutational processes involved in the evolution of homologous clusters are unlikely to occur evenly across entire genomes, and will almost certainly be subject to local effects, which will complicate matters. As an illustration of the problems inherent in determining mutational biases, consider gene loss. Developmental genes are thought to be unusual in that relatively simple mutations, either in the gene itself or in cis-regulatory elements, can provide new functions for otherwise redundant duplicated genes. Many developmen-

tal proteins have a modular structure, which allows temporal or spatial expression patterns to be altered by relatively small mutational changes, while the evolutionarily conserved core functions are retained.[23] In contrast, metabolic genes have very limited possibilities of gaining a new function prior to loss by mutational drift due to their rigid functional requirements. Thus the relative likelihood of gene loss following duplication will depend on the class of the duplicated gene, as well as other factors such as the extent to which there is the potential and the need for further adaptation of the gene in question.

Could the use of gene order data help us to understand the evolution of homologous clusters? If gene order is conserved between homologous clusters then the probability that the homologous clusters arose by chance becomes even more remote. Such an approach has been used by Pebusque et al.[24] to infer a quadruplicated region on human chromosomes 4, 5, 8, and 10. They started from a region on chromosome 8 for which they possessed detailed mapping information, then searched for paralogs, and finally deduced duplication events from mapping and phylogenetic data. Such a study suggests that duplicated regions may be common in mammalian genomes, and support reports of extensive regional duplications (see Reference 25), although the criteria commonly used to infer paralogy are perhaps too broad.[26]

The adaptive hypothesis vs. the block duplication hypotheses

Can gene order help to decide between the adaptive and block duplication hypotheses? It appears not, since the two hypotheses cannot be readily discriminated on the basis of linkage patterns of related clusters. Local gene shuffling by inversions can upset the initially conserved gene order of block duplicated clusters (as in yeast[27]), and it is not clear whether adaptive gene clustering is likely to imply adaptive gene ordering. In the case of the homologous clusters described above, gene order appears to be weakly conserved between the clusters on human chromosomes 6 and 9.[6]

The preponderance of homologous clusters may well affect the likelihood of the alternative hypotheses. The existence in yeast of 55 duplicated regions extending across half the genome[27] strongly suggests that a block duplication event is responsible. Two problems suggest that the adaptive hypothesis is unlikely to explain so many duplicated regions. First, it may be reasonable to suppose that selection might favour physical linkage of some genes, but is such selection likely for a high proportion of all genes? Second, has enough time elapsed for sufficient mutational events to have caused a thorough reshuffling of the entire genome?

Further suggestive evidence might be obtained from consideration of the intronic content of genes within related clusters. If the multiple copies of a gene have arisen by block duplication, then one would expect similar intronic contents. However, if multiple copies have been spread around the genome by retrotransposition, as is consistent with the adaptive hypothesis, then one copy might contain introns whereas the other copies are intronless. Therefore, a preponderance of intronless genes would provide suggestive evidence in favour of the adaptive hypothesis, unless it could be demonstrated that the original, later to be duplicated, cluster contained intronless genes, in which case all copies should be intronless. This proposed test is unlikely to be conclusive because retrotransposition is not the only mode of gene movement, and because subsequent intron evolution may well have obscured any intron patterns, which might have been generated during homologous cluster evolution.

Polyploidisation or local block duplications?

If the block duplication hypothesis for the evolution of homologous clusters is favoured over the adaptive hypothesis there remains the question of whether the block duplications were small and independent, or whether they involved the entire genome. One approach would be to consider the estimated time of duplication for a number of homologous clusters. The timing of a duplication could be estimated from the shape of the tree, by comparing the duplication event with the appearance of various groups. This method is dependent on the accuracy of tree reconstruction.

In the case of yeast, Wolfe and Shields[27] provided two reasons to believe that a polyploidisation event, rather than many independent and small block duplications, was responsible for the large number of duplicated regions. First, the orientation of both homologous clusters with respect to the centromere was the same in a significantly high proportion of cases. This suggests that the duplications were not independent. Second, no triplicated regions were found, which argues against a series of duplications spread over time.

Conclusions

Until we have the complete genomes of a number of vertebrates, and in particular those of the early chordates, it will be difficult to evaluate general hypotheses for the evolution of homologous clusters. When we possess such information, we shall be able to address a number of questions. How common are homologous clusters within different genomes? Are there more homologous clusters than we would have expected by chance? What sort of genes does one find within the homologous clusters? Is gene order, and also orientation with respect to the centromere, conserved between homologous clusters? And finally, did these homologous clusters all arise at the same time, at some ancestral vertebrate polyploidisation event?

In the meantime (until about 2005 for the human genome[28]), we are restricted to the level of anecdote, and we should not infer from a few case studies assumptions about

the entire genome. One way in which case studies may prove profitable is in examining the hypothesis that homologous clusters may be adaptive. A fuller understanding of the regulation of genes within clusters might show how regulation is dependent on physical proximity, or perhaps even the relative position and orientations of a number of genes. The regulation of transgenic rearranged homologous clusters would then be different from the wild type, and a transgenic phenotype would provide concrete evidence that selection not only cares about the sequence of a gene but also its context with respect to other genes.

Acknowledgments

The authors thank Peter Holland, Seb Shimeld, L. Lundin, Adam Wilkins, and two anonymous referees.

References

1. Lundin LG. Evolution of the vertebrate genome as reflected in paralogous chromosomal regions in man and the house mouse. Genomics 1993;16: 1–9.
2. Ruddle FH, Bentley KL, Murtha MT, Risch N. Gene loss and gain in the evolution of the vertebrates. Development 1994;155–161.
3. Sharman AC, Holland PWH. Conservation, duplication, and divergence of developmental genes during chordate evolution. Netherlands J Zool 1996;46:47–67.
4. Bailey WJ, Kim J, Wagner GP, Ruddle FH. Phylogenetic reconstruction of vertebrate Hox cluster duplications. Mol Biol Evol 1997;14:843–853.
5. Hughes AL. Phylogenetic tests of the hypothesis of block duplication of homologous genes on human chromosomes 6, 9, and 1. Mol Biol Evol 1998;15:854–870.
6. Endo T, Imanishi T, Gojobori T, Inoko H. Evolutionary significance of intra-genome duplications on human chromosomes. Gene 1997;205: 19–27.
7. Kasahara M, Hayashi M, Tanaka K, Inoko H, Sugaya K, Ikemura T, Ishibashi T. Chromosomal localization of the proteasome Z subunit gene reveals an ancient chromosomal duplication involving the major histocompatibility complex. Proc Natl Acad Sci USA 1996;93:9096–9101.
8. Katsanis N, Fitzgibbon J, Fisher EMC. Paralogy mapping: identification of a region in the human MHC triplicated onto human chromosomes 1 and 9 allows the prediction and isolation of novel PBX and NOTCH loci. Genomics 1996;35:101–108.
9. Patton SJ, Luke GN, Holland PWH. Complex history of a chromosomal paralogy region: Insights from amphioxus aromatic amino acid hydroxylase genes and insulin-related genes. Mol Biol Evol 1998;15:1373–1380.
10. Nadeau JH, Sankoff D. Comparable rates of gene loss and functional divergence after genome duplications early in vertebrate evolution. Genetics 1997;147:1259–1266.
11. Wagner A. The fate of duplicated genes: loss or new function? Bioessays 1998;20:785–788.
12. Wilkins AS. Canalization: a molecular genetic perspective. Bioessays 1997;19:257–262.
13. Cooke J, Nowak MA, Boerlijst M, Maynard-Smith J. Evolutionary origins and maintenance of redundant gene expression during metazoan development. Trends Genet 1997;13:360–364.
14. Holland PWH. Vertebrate evolution: something fishy about Hox genes. Current Biology 1997;7:R570–R572.
15. Hurst LD. The evolution of genomic anatomy. TREE 1998; 14:108–112.
16. Blumenthal T. Gene clusters and polycistronic transcription in eukaryotes. Bioessays 1998;20:480–487.
17. Lyttle TW. Segregation distorters. Ann Rev Genet 1991;25:511–557.
18. Ferris PJ, Goodenough UW. The mating-type locus of *Chlamydomonas reinhardtii* contains highly rearranged DNA sequences. Cell 1994;76:1135–1145.
19. Felsenstein J. Cases in which parsimony or compatibility methods will be positively misleading. Syst Zool 1978;27:401–410.
20. Philippe H, Laurent J. How good are deep phylogenetic trees? Curr Opin Genet Dev 1998;8:616–623.
21. Ting CT, Tsaur SC, Wu ML, Wu CI. A rapidly evolving homeobox at the site of a hybrid sterility gene. Science 1998;282:1501–1504.
22. Nadeau JH, Sankoff D. Counting on comparative maps. Trends Genet 1998;14:495–501.
23. Duboule D, Wilkins AS. The evolution of 'bricolage'. Trends Genet 1998;14:54–59.
24. Pebusque M-J, Coulier F, Birnbaum D, Pontarotti P. Ancient large-scale genome duplications: phylogenetic and linkage analyses shed light on chordate genome evolution. Mol Biol Evol 1998;15:1145–1159.
25. http://www.cib.nig.ac.jp/dda/timanish/poster/poster.html.
26. Skrabanek L, Wolfe KH. Eukaryote genome duplication—where's the evidence? Curr Opin Genet Dev 1998;8:694–700.
27. Wolfe KH, Shields DC. Molecular evidence for an ancient duplication of the entire yeast genome. Nature 1997;387:708–713.
28. Rowen L, Mahairas G, Hood L. Sequencing the human genome. Science 1997;278:605–607.

chromosomes is inactivated in female soma, thus in the individual cells of both male and female mammals only one X chromosome is expressed. However, not all X-linked genes are inactivated (Lahn and Page 1997), and these genes often have functional Y homologs (Ellis 1998).

Genomic imprinting

Genomic imprinting arises as a combination of epigenetic inheritance and parent-of-origin effects, and is manifested by the expression of a particular gene in a particular cell limited to the allele derived from a particular parent (Efstratiadis 1994). Imprinting has been well documented in mammals but has also been demonstrated in a variety of other eukaryotes: maize, zebra fish, dipterans and *Drosophila melanogaster* (Lloyd *et al.* 1999). In many of these organisms gene silencing is associated with heterochromatisation but not necessarily methylation status (Lloyd *et al.* 1999). It is unclear whether the known examples of genomic imprinting represent one ancient and conserved process or a number of similar independently evolved mechanisms (Lloyd *et al.* 1999), although the diversity of mechanisms appears to suggest the latter (Brannan and Bartolomei 1999).

Usually genes are unidirectionally imprinted, but *GNAS1* is bidirectionally imprinted in that it encodes maternally, paternally and biallelically expressed products (Hayward 1998). The evolution of imprinting presents a paradox from the point of view of ploidy level evolution. Imprinting is a hybrid between haploidy and diploidy, but appears to incur the costs of both ploidy levels without enjoying the benefits of either: "in some ways imprinting is the worst of both ploidy worlds" (Spencer and Williams 1997). How can we explain this contradiction of Panglossian thinking?

We need to explain numerous phenomena associated with genomic imprinting (for a review see Hurst 1997), and we can use either evolutionary or mechanistic arguments. It is important to note that evolutionary and mechanistic theories of imprinting are not in competition. However, some mechanistic features might give us clues about the evolutionary origins of imprinting. For instance, the epigenetic inheritance of imprinting is thought to be controlled by methylation patterns which induce changes in histone acetylation patterns, chromatin structure, and transcriptional activity (Bestor 1998; Constancia *et al.* 1998; Jones 1999; Kass *et al.* 1997; Wutz *et al.* 1997). Furthermore, imprinted genes are often found to contain repeated units (Neumann *et al.* 1995; Shibata *et al.* 1998). When put together with the phenomenon of cosuppression of repeated elements (Garrick *et al.* 1998; Jensen *et al.* 1999; Kricker *et al.* 1992), these findings offer a mechanistic explanation for the origins of imprinting (Barlow 1993).

There exist a number of evolutionary explanations for imprinting, which can be divided into three general classes (see Hurst 1997). The first class of explanation is that imprinting is maintained as a side product of alternative processes (as suggested above for the origins of

imprinting). Secondly, imprinting may be adaptive for the organism. Imprinting may provide a defence mechanism against parthenogenesis, ovarian trophoblast disease, invasive placentas, chromosome loss or gain, or dominant deleterious somatic mutations; alternatively imprinting may improve gene regulation by minimising expression variance, controlling cellular differentiation, or allowing temporal control.

The third theory is the conflict theory of imprinting (Moore and Haig 1991), which is an extension of the classical theory of parent-offspring conflict (Trivers 1974). As originally conceived (Moore and Haig 1991), imprinting derives from multiple paternity. All offspring within a brood are equally related via maternal inheritance, but may be unrelated via paternal inheritance. Hence maternally derived genes will favour a more equitable allocation of resources than paternally derived genes. Genomic imprinting allows parent-of-origin specific effects, with paternally expressed genes enhancing growth and maternally expressed genes suppressing growth.

Haig (1997) has broadened the scope of the conflict theory to suggest that all interactions among relatives can potentially lead to imprinting. Under Haig's ESS formulation genomic imprinting is favoured for genes with parentally antagonistic effects which are "associated with an inclusive fitness benefit when derived from one parent but an inclusive fitness cost when derived from the other" (Haig 1997). Burt and Trivers (1998) have pointed out that conflicts also exist between the different genetic components of imprinting as well as between individuals.

Although there exist differences between the various explicit models of imprinting (Hurst and McVean 1998; Spencer *et al.* 1999), some versions of the conflict theory provide testable predictions. The conflict theory correctly predicts that imprinted genes should affect growth, but this is not a discriminating prediction. More explicitly, the conflict theory suggests that maternal imprints should suppress growth while paternal imprints should enhance growth, a prediction which holds in the case of *Igf2* and *Igf2r* (Haig and Graham 1991). However, neither the mouse knockout data (Hurst and McVean 1998) nor the uniparental disomy data (Hurst and McVean 1997) provide conclusive support for the ability of the conflict hypothesis to predict the covariance of the direction and effect of an imprinted gene. However, Haig (1997) has claimed that the knockout data do not necessarily contradict the conflict hypothesis, and *post hoc* arguments might explain the uniparental disomy data (Hurst and McVean 1998). It is not clear whether the conflict hypothesis predicts that all genes affecting growth should be imprinted: for example, should *Igf1* be imprinted? Haig (1997) has suggested that imprinting is not expected if genes are insufficiently parentally antagonistic, perhaps if gene dosage sensitivity is low. An alternative explanation proposes that some genes may not be imprinted because of the effects of deleterious mutations (Mochizuki *et al.* 1996).

The prediction of the conflict hypothesis that imprinting should not persist into adulthood is broadly supported, although exceptions to the rule are accumulating (Hurst and McVean 1998). Although not all models of genomic imprinting suggest that multiple paternity is strictly necessary

for imprinting to evolve, it is agreed that imprinting is favoured by multiple paternity (Hurst and McVean 1998). The finding that some monogamous rodents display imprinting is thus not predicted by the conflict theory, although monogamy may have evolved only recently (Hurst 1998; Vrana *et al.* 1998).

Conflict is thought to lead to antagonistic coevolution, also known as arms races, which should cause rapid evolution at interacting sites (McVean and Hurst 1997b). The prediction of conflict leading to high rates of evolution appears to hold for host-parasite and maternal-foetal interactions, but imprinted genes do not appear to be rapidly evolving (McVean and Hurst 1997b). The prediction that interacting regions should be particularly rapidly evolving was tested by examining the molecular evolution of *Igf2* and *Igf2r* in mammals. The two genes are imprinted in the directions predicted by the conflict theory. The protein Igf2 is a paternally expressed growth factor, and the protein Igf2r is a maternally expressed multifunctional receptor protein which binds Igf2 and transports it to the lysosomes to be destroyed. If ongoing conflict is occurring within the regions of interaction between Igf2 and Igf2r, then the presence of positive selection should be indicated by elevated rates of molecular evolution within such regions. In a sliding window comparison between mouse and rat, Igf2r is evolving slowly at the site of interaction with Igf2 (McVean and Hurst 1997b). Furthermore in comparisons of nine mammalian *Igf2* sequences no nonsynonymous substitutions were found to affect the site of binding to Igf2r (McVean and Hurst 1997b).

It was not clear from this initial study of *Igf2r* whether the observed result of a low rate of molecular evolution at the Igf2 binding site was simply due to chance. To test whether the patterns across the gene were due to stochastic processes I compared the intragenic *Igf2r* sliding window plots of the mouse versus rat comparison against the human versus cow comparison (see Research Paper 4 for details). For both comparisons the Igf2 binding site was slowly evolving, and furthermore the evolutionary rate plots of the two independent pairwise species comparisons were found to be significantly similar, from which it could be inferred that deterministic rather than stochastic processes were influencing rates of evolution. These results can be taken as evidence against the conflict theory prediction of antagonistic coevolution (as in Burt and Trivers 1998), although it has been suggested that imprinting may lead to a stable ESS (Haig 1997).

Although I found no evidence for antagonistic coevolution at the Igf2 binding site, there were a couple of regions within the gene which were evolving rapidly. The use of PAM matrices to classify the amino acid changes indicated that chemically different amino acids were being substituted within one of these regions, the signal sequence which is involved in determining the intracellular location of the protein. The signal sequence of *Igf2r* was also shown to be evolving faster than a small sample of signal sequences from other genes. This evidence is consistent with positive selection acting on the *Igf2r* signal sequence, which suggested that antagonistic conflict might be occurring, but at the signal sequence rather than the Igf2 binding site. Perhaps by changing its signal sequence the cellular location of Igf2r is altered so as to better sequester Igf2.

The unusual signal sequence evolution of *Igf2r* has provided the first evidence of unusual molecular evolution in imprinted genes.

Another issue that is pertinent to the evolution of imprinting is the question of whether imprinted genes are functional. If the imprinted gene is non-functional, then one would not expect the potentially costly process of imprinting to be maintained, and one would also expect the gene to undergo mutational decay. In mammals *H19* is genomically imprinted and appears to be a RNA-coding gene (Bartolomei 1997), although whether the spliced transcript has a function is uncertain. I have shown that the spliced transcript of *H19* has been maintained by stabilising selection. When the mouse and rat sequences are compared the introns of *H19* have evolved at 2.5 times the rate of the exons, which indicates strong stabilising selection on the exons since this ratio is significantly higher than the ratios of a reference set of 41 mouse-rat orthologous protein-coding genes. The difference between intronic and exonic rates does not appear to be due to within gene variation in the mutation rate since all four introns evolve faster than the surrounding exons (see Research Paper 5 for further details).

The tendency of imprinted genes to cluster (Reik and Maher 1997) is clearly pertinent to my thesis of genomic anatomy. A number of explanations have been proposed for this linkage effect. Burt and Trivers (1998) suggest that clustering may come about as a response to conflict between different genes affecting the imprinting process. Alternatively, Haig (1997) has suggested a mutational constraint, and Mochizuki *et al.* (1996) have proposed that clusters of imprinted genes are favoured because the imprinting process is costly.

X-linked imprinting

The two monoallelic expression mechanisms of X linkage and genomic imprinting come together in the phenomenon of X-linked imprinting. Although the conflict hypothesis remains the most attractive hypothesis for autosomal genomic imprinting (Hurst and McVean 1998), it appears that cooperation rather than conflict may underpin X-linked imprinting (Pagel 1999). Under the cooperation hypothesis, mothers and fathers agree about the inheritance of sex specific traits which favour their offspring, and one way of solving the problems of sex chromosome dosage compensation and sex limited traits is through X-linked imprinting (Iwasa 1998). The transmission of traits beneficial to females is enabled by the epigenetic inheritance of the paternal X chromosome only. The paternal X is always inherited by daughters, and was passed down from a mother in the previous generation. The maternal X chromosomes are inherited by both sons and daughters, and may have been passed down from either a mother or a father in the previous generation. The cooperation theory explains why mice lacking the paternal X chromosome are larger than mice lacking the maternal X chromosome (Thornhill and Burgoyne 1993), a result contrary to the predictions of the conflict hypothesis (Hurst 1997).

The evolutionary consequences of hemizygous expression

Imprinted genes are monoallelically expressed but biallelically transmitted. Similarly, X-linked genes are monoallelically expressed in males but biallelically transmitted through females. So both X-linked and imprinted genes combine haploidy and diploidy. What are the consequences of hemizygous expression?

X-linked genes undergo haploid selection in males and diploid selection in females (assuming random X chromosome inactivation in females). Thus the mutation load of X-linked genes is the same as that of all genes in haplodiploids, that is $L \sim 3\mu/2$ (Hedrick and Parker 1997). When in heterozygotes, imprinted genes are hemizygously selected as the wildtype allele half the time and as the mutant allele the other half of the time. This means that imprinted genes are subject to the same selection as biallelically expressed genes in a diploid with no dominance (h=1/2). Indeed, the phenomenon of imprinting has been considered from the view of dominance modification (Spencer *et al.* 1999). Since the result of $L \sim 2\mu$ in diploids is highly insensitive to the level of dominance (Crow and Kimura 1970), the mutational load of imprinted genes in diploids is the same as that of biallelically expressed genes in diploids.

Charlesworth *et al.* (1987) have analysed how substitution rates might differ between X-linked and autosomal genes. When sex ratios are unbiased then the X chromosome has a higher substitution rate of both recessive advantageous and dominant deleterious mutations. When the operational sex ratio is female biased then the higher substitution rate remains for recessive advantageous mutations remains but disappears for dominant deleterious mutations (McVean 1997). As for imprinted genes, their rate of substitution relative to biallelically expressed autosomal genes would appear to depend on the dominance of the mutations in autosomal genes. If h>1/2 then advantageous genes substitute faster and deleterious genes slower on the autosomes, and if h<1/2 then deleterious genes substitute faster and advantageous genes slower on the autosomes.

The evolution of mutation rates

What factors explain the evolution of mutation rates? The argument from constraints notes that mutation rates are very low and perhaps cannot be further reduced (Sturtevant 1937). But the fact that different species possess widely different mutation rates (Drake 1991) argues against the suggestion that mutation rates are as low as they can possibly be (Dawson 1998). More conclusively, Nothel (1987) has demonstrated the evolutionary adjustment of mutation rates in *Drosophila*.

Two hypotheses involving trade-offs are thought to account for adaptive mutation rates (Dawson 1998; Sniegowski 1997). The "minimal" (sensu Maynard Smith 1978) mutation rate

argument supposes that reducing the mutation rate imposes costs just as having a high mutation rate imposes costs from deleterious mutations. Alternatively, the "optimal" (sensu Maynard Smith 1978) mutation rate comes from considerations of long term adaptation, a balance between the deleterious mutation load and the ability to adapt by advantageous mutations (Kimura 1967; Leigh 1970; Leigh 1973).

In an asexual population the "optimal" mutation rate is greater than zero because the lineage containing the modifier of zero mutation rate is doomed to extinction (as originally argued by Fisher 1930). The production of enough advantageous mutations to compensate for the majority of mutations being deleterious is what allows "mutator strains" of bacteria to spread (Chao and Cox 1983; Sniegowski *et al.* 1997). In a sexual population the "optimal" mutation rate is zero because the modifier can recombine into organisms with advantageous mutations. The effect of a "minimal" mutation rate explains why a zero mutation rate would not be adaptive in sexual organisms (Leigh 1970).

McVean and Hurst (1997a) have extended Leigh's (1970) model to account for unlinked modifiers of both X-linked and autosomal mutation rates in a diploid sexual population (for details see McVean 1997). Under a simple analytical model invoking only deleterious mutations, and ignoring sex biases in the mutation rate, the relative advantage of an X-linked modifier versus an autosomal modifier can be calculated in terms of the mutation rate, the strength of selection against mutations, and the dominance. This analytical model suggests that modifiers of the X-linked mutation rate are more strongly favoured than autosomal mutation rate modifiers when dominance is low ($h<1/2$), a qualitative conclusion which is reinforced by more realistic simulation models (McVean 1997).

It appears that this result can be extended to the case of imprinted genes simply by inspection of the appropriate equations (see Page 136 in McVean 1997; see Box 1 in McVean and Hurst 1997a). As argued above, the evolution of imprinted genes is the same as that of biallelically expressed genes with no dominance ($h=1/2$). Since the advantage to an autosomal modifier is proportional to the level of dominance, then as long as $h<1/2$ for biallelically expressed autosomal genes then the modifier of imprinted mutation rate is more favoured.

The above analysis is complicated by sex biases in the mutation rate and biases in the sex ratio. McVean and Hurst (1997a) showed that, as long as the male to female mutation rate bias was no greater than about 5, then the modifiers of mutation rate should still be more favoured on the X chromosome. A male bias in mutation rates effectively reduces the X-linked mutation rate versus the autosomal mutation rate since X chromosomes spend 1/3 of their time in males and autosomes spend 1/2 of their time in males (assuming an unbiased sex ratio). This means that the relative advantage of a reduction in the mutation rate of the X chromosome is reduced.

McVean and Hurst (1997a) tested their prediction that the mutation rate of the X chromosome should be lower than the mutation rate of the autosomes by sequence comparison of

mouse and rat sequences. Their finding of X-linked K_S significantly lower than autosomal K_S can be taken in favour of the notion that mutation rates are adaptive (McVean and Hurst 1997a).

I have extended the analysis of Hurst and McVean (1997a) to consider imprinted genes as well as X-linked and autosomal genes (see Research Paper 6 for details). The finding that both imprinted and X-linked genes have significantly lower K_S than autosomal genes suggests that their mutation rates are adaptive.

Moreover, this analysis hints at a possible limit to the efficacy of selection on the mutation rate, which is relevant to my thesis since the limitations of selection mean that not all genomic anatomy need make sense. The greater the number of genes under modification the greater the strength of selection on the modifier. The mutation rate appears to be modified on a chromosomal basis, as shown by the X chromosome, and on a regional basis, as shown by the fact that imprinted genes tend to cluster. But those imprinted genes which are not found in clusters do not have low mutation rates, which suggests that selection on modifiers of the mutation rate reaches its limit at the single gene. This limit to selection appears to concur with Wright's (1934) arguments against Fisher's (1928) theory of dominance modification. A modifier of the mutation rate of a single gene can only evolve at a rate of the order of the mutation rate of a single gene. Furthermore a costless modifier may not be possible, and the modifier might suffer mutation pressures itself.

Research Paper 4. The molecular evolution of *IGF2R*

Nick Smith and Laurence Hurst (1998)

Genetics **150** 823-833.

Molecular Evolution of an Imprinted Gene: Repeatability of Patterns of Evolution Within the Mammalian Insulin-Like Growth Factor Type II Receptor

Nick G. C. Smith and Laurence D. Hurst

Centre for Mathematical Biology, School of Biology and Biochemistry, University of Bath, Bath BA2 7AY, United Kingdom

Manuscript received March 2, 1998
Accepted for publication June 30, 1998

ABSTRACT

The repeatability of patterns of variation in K_a/K_s and K_s is expected if such patterns are the result of deterministic forces. We have contrasted the molecular evolution of the mammalian insulin-like growth factor type II receptor (*Igf2r*) in the mouse-rat comparison with that in the human-cow comparison. In so doing, we investigate explanations for both the evolution of genomic imprinting and for K_s variation (and hence putatively for mutation rate evolution). Previous analysis of *Igf2r*, in the mouse-rat comparison, found K_a/K_s patterns that were suggested to be contrary to those expected under the conflict theory of imprinting. We find that K_a/K_s variation is repeatable and hence confirm these patterns. However, we also find that the molecular evolution of *Igf2r* signal sequences suggests that positive selection, and hence conflict, may be affecting this region. The variation in K_s across *Igf2r* is also repeatable. To the best of our knowledge this is the first demonstration of such repeatability. We consider three explanations for the variation in K_s across the gene: (1) that it is the result of mutational biases, (2) that it is the result of selection on the mutation rate, and (3) that it is the product of selection on codon usage. Explanations 2 and 3 predict a K_a–K_s correlation, which is not found. Explanation 3 also predicts a negative correlation between codon bias and K_s, which is also not found. However, in support of explanation 1 we do find that in rodents the rate of silent C ⇒ T mutations at CpG sites does covary with K_s, suggesting that methylation-induced mutational patterns can explain some of the variation in K_s. We find evidence to suggest that this CpG effect is due to both variation in CpG density, and to variation in the frequency with which CpGs mutate. Interestingly, however, a GC4 analysis shows no covariance with K_s, suggesting that to eliminate methyl-associated effects CpG rates themselves must be analyzed. These results suggest that, in contrast to previous studies of intragenic variation, K_s patterns are not simply caused by the same forces responsible for K_a/K_s correlations.

IN mammals there exists much variation in synonymous substitution rates (K_s) and the ratio of nonsynonymous to synonymous substitution rates (K_a/K_s), both within (ALVAREZ-VALIN *et al.* 1998) and between (WOLFE and SHARP 1993) genes. There are two alternative explanations for such variation: that it is simply a consequence of chance (*e.g.*, some genes, or parts of them, might just happen to mutate faster than others) or that it is due to deterministic effects. Possible deterministic forces include selection and repeatable mutational biases. Analysis of the repeatability of patterns of molecular evolution in independent comparisons is one means to distinguish between stochastic and deterministic explanations (MOUCHIROUD *et al.* 1995). Here we ask about the repeatability of patterns of intragene variation in K_a/K_s and K_s. To the best of our knowledge this is the first example where repeatability of intragene variation in both of these parameters has been examined. The gene that we examine, the insulin-like growth factor type II receptor (*Igf2r*), is of interest as it is one of a small class of mammalian genes that are imprinted. It has also recently been implicated in the determination of IQ (CHORNEY *et al.* 1998).

Genomically imprinted genes are those for which expression is dependent upon the sex of the parent from which they are derived (EFSTRATIADIS 1994). Well-described examples include murine insulin-like growth factor II (*Igf2*), a growth factor expressed from the paternally derived genome in fetuses, and the murine insulin-like growth factor type II receptor (*Igf2r*) expressed from the maternally derived genome. In eutherians and marsupials, but not in birds (ZHOU *et al.* 1995), the protein product of this gene binds to Igf2 and takes it to the lysosome for digestion. The gene is also known as the cation-independent mannose-6-phosphate (M6P) receptor as it is also known to bind M6P (KORNFELD 1992). The latent complex of transforming growth factor beta (TGFβ) also binds to the Igf2r protein via the M6P binding domains (DENNIS and RIFKIN 1991). Additionally the Igf2r protein has the ability to bind retinoic acid (KANG *et al.* 1998). The binding activities of Igf2

Corresponding author: Nick Smith, Centre for Mathematical Biology, School of Biology and Biochemistry, University of Bath, Claverton Down, Bath BA2 7AY, UK. E-mail: n.smith@bath.ac.uk

and M6P are known to occur at different sites within the Igf2r protein. In rats *Igf2r* is known to be imprinted in the liver (MILLS *et al.* 1998) and in humans *IGF2R* appears to be imprinted in some but not all individuals (OGAWA *et al.* 1993; XU *et al.* 1993). We are unaware of data pertinent to the imprinting status of the gene in cow. However, from our knowledge that the mouse, rat, and human *IGF2R* genes are now imprinted, we can be sure that some of the changes revealed by human-cow and mouse-rat comparisons of *Igf2r* will have occurred in an imprinted gene.

Aside from allowing us to ask whether patterns of K_a/K_s and of K_s within genes are repeatable and hence deterministically based, the patterns of molecular evolution in this gene are of interest for two reasons. First, the pattern of variation in K_s across the gene may shed light on the evolution of mutation rates or of codon usage patterns. Second, the variation in K_a/K_s is potentially informative as to the evolution of imprinting. One theory for the evolution of imprinting would find discriminating support from evidence for antagonistic coevolution between imprinted genes (see MCVEAN and HURST 1997b; TRIVERS and BURT 1998; but see also HAIG 1997).

K_s variation, mutation rate evolution, and imprinted genes: We consider four hypotheses to explain variation in K_s across *Igf2r*.

Hypothesis 1: K_s variation may be caused by stochastic forces, in which case K_s will not be repeatable. The remaining three hypotheses suppose that K_s variation is the result of deterministic forces that would cause K_s repeatability.

Hypothesis 2: If silent sites are considered neutral then K_s is an unbiased estimator of the mutation rate (KIMURA 1983), and so K_s variation may be the result of deterministic mutational processes, with mutational biases dependent on perhaps sequence composition or regional effects. The two final hypotheses, (3) and (4), both suggest that K_s variation is molded by selection, but differ with respect to the character that selection is acting upon.

Hypothesis 3: Decreasing mutation rates is expensive both in time and energy, while the benefits of a lower mutation rate depend on the functional importance and mutational sensitivity of the region concerned. Such reasoning has been used to explain the constant per genome mutation rate across many unicellular organisms (DRAKE 1991). If this trade-off theory of mutation rate evolution is correct, and if selection is strong enough, one might imagine that selection would favor imprinted genes to have low mutation rates. This prediction derives from imprinted genes having haploid expression and hence recessive mutations in imprinted genes being unmasked. Hence it may be optimal for an organism to invest more into reducing the mutation rate of imprinted genes below the mutation rate of comparable nonimprinted ones. It should, however, be noted that selection on the modifier is likely to be very weak.

From the same logic, one might expect that the most important parts of an imprinted gene should be provided with especially low mutation rates. Thus if silent sites in the mammalian *Igf2r* gene are neutral, then the variation in K_s (and hence mutation rate) across the gene may be the result of selection acting to optimize mutation rates. We term this explanation of K_s variation the "selected mutation rate" hypothesis.

Hypothesis 4: ALVAREZ-VARIN *et al.* (1998) have found intragenic correlations between synonymous and nonsynonymous substitution rates in mammalian genes. They conclude that these correlations might well be due to common selective constraints, such as selection for translational accuracy, between synonymous and nonsynonymous sites. This argument suggests that K_s in mammals is not the mutation rate (as often assumed), but instead reflects the local strength of selection. We term this explanation of K_s variation the "common constraints" hypothesis.

Just as for the selected mutation rate argument, the common constraints hypothesis would predict that imprinted genes, due to their haploid expression, would be under greater selective pressures, and thus should have lower K_s values than nonimprinted genes. The empirical data on this point are equivocal. Previous analyses have found *Igf2* to have one of the lowest K_s values in the rodent genome (MCVEAN and HURST 1997b). Imprinted genes as a class have a K_s ($K_s = 19.87$, $N = 7$) between that of autosomes ($K_s = 22.9$, $N = 238$) and *X*-linked genes ($K_s = 14.63$, $N = 33$; MCVEAN and HURST 1997a,b). However, on nonparametric analysis the imprinted genes are found to be not significantly different from autosomal genes. Given sample size limitations, however, the issue cannot be considered fully resolved.

K_a/K_s variation and the evolution of genomic imprinting: Numerous theories have been proposed for the evolution of genomic imprinting (for a review see HURST 1997). The "conflict" hypothesis (MOORE and HAIG 1991) proposes that imprinting is an intraindividual manifestation of classic parent-offspring conflict (TRIVERS 1974). According to this model, paternally derived genes are under selection to extract resources from mothers whereas maternally expressed genes are under selection to oppose this (for models see MOCHIZUKI *et al.* 1996; SPENCER *et al.* 1998). This hypothesis makes predictions about the growth effects of paternally and maternally expressed genes and it is presently unclear whether these predictions are borne out (HURST and MCVEAN 1997).

Given that maternal and paternal genes have antagonistic interests, the conflict theory is unique in predicting that an arms race may develop between maternally and paternally derived genes for the control of fetal growth demands. Such antagonistic coevolution may be mediated through changes in the structure of

the proteins concerned. Were one to find evidence for rapid evolution of imprinted genes (*i.e.*, a high K_a/K_s ratio) this, it has been argued, could then reasonably be taken as evidence favoring the conflict hypothesis (McVEAN and HURST 1997b; see also TRIVERS and BURT 1998). Unfortunately this is not a falsifying prediction as the finding of slow evolution need not be considered evidence against the hypothesis (HAIG 1997; McVEAN and HURST 1997b) for a variety of reasons.

Comparable maternal-fetal conflict is a good candidate explanation for the rapid evolution of numerous genes, such as the placental lactogens (WALLIS 1993), prolactin (WALLIS 1981), the homeobox gene *Pem* (SUTTON and WILKINSON 1997), growth hormones (WALLIS 1994), the pregnancy-associated glycoproteins (XIE *et al.* 1997), and numerous others (see McVEAN and HURST 1997b). Antagonistic coevolution is similarly the most likely explanation for the rapid changes seen in antigenic components of parasites and antigen recognition components of immune systems (for example see HUGHES *et al.* 1994), and may be invoked to explain rapid changes in a gene putatively involved in sexual conflict (TSAUR and WU 1997).

Perhaps surprisingly then, an analysis of the molecular evolution of seven imprinted genes revealed them to be no faster evolving at the protein level (controlling for K_s, and hence possibly for mutation rate, variation) than ordinary receptors and significantly slower evolving than immune system genes (McVEAN and HURST 1997b). In the same study, an analysis of molecular evolution within *Igf2r*, compared between mouse and rat, showed that where Igf2r binds to Igf2 a low K_a/K_s ratio is found [indicating, contrary to initial expectations (McVEAN and HURST 1997b), stabilizing rather than disruptive selection]. This analysis also shows two high peaks in K_a/K_s that were not commented on. One of these was at the signal sequence and one at the position at which Igf2r binds M6P.

Neither of these peaks in the pattern of protein evolution rates are expected. Signal sequences direct protein transport within the cell, and possess the conserved structure of a hydrophobic core flanked by a polar basic region and a hydrophilic region (KENDREW 1994). Since artificial signal sequence constructs lacking such features fail to perform their normal functions (IZARD *et al.* 1996), one would expect strong amino acid conservation at signal sequences and hence low K_a/K_s values. Igf2r assists the transport of acid hydrolases to the lysosomes by binding the enzymes at their phosphomannosyl residues using the M6P binding site (KORNFELD 1992). Because the phosphomannosyl residues are high-affinity ligands there should be high sequence specificity at the M6P binding site, and hence strong amino acid conservation and low K_a/K_s values.

Here we ask whether the peaks and troughs in the K_a/K_s pattern across the gene revealed in the mouse-rat comparison are also observed in the human-cow comparison or whether they are statistical artifacts and hence not worth further investigation. That the mouse-rat patterns might be artifacts is possible since the sliding window used was small and hence the expected error in estimates of K_a/K_s per window is high.

MATERIALS AND METHODS

Extraction and analysis of genes: NCBI Entrez (http://www.ncbi.nlm.nih.gov/Entrez/) was used to search for complete coding mammalian *Igf2r* sequences. Four were obtained: human, cow, mouse, and rat (accession numbers J03528, J03527, U04710, and U59809, respectively). The orthologies of these sequences were confirmed using BLASTN (ALTSCHUL *et al.* 1990) and the HOVERGEN database (DURET *et al.* 1994).

Sequence manipulation and alignment was performed using programs in the Wisconsin package (GENETICS COMPUTER GROUP 1994). All four mammalian Igf2r protein-coding DNA sequences were aligned using the default DNA alignment parameters of the program PILEUP. The program GAPFRAME was used to move gaps to codon boundaries, and the resultant alignment was very similar to the protein alignment obtained using the default protein alignment parameters of PILEUP. All four gapped sequences were then cut into 24 orthologous nonoverlapping sections each of 300 bp, or 74 orthologous overlapping sections (300 bp every 100 bp). Synonymous and nonsynonymous nucleotide substitution rates (K_s and K_a) were estimated for all mouse-rat and human-cow orthologous pairs using the program KESTIM (COMERON 1995) with KIMURA's (1980) two-parameter model for multiple hits correction. The mouse-rat and human-cow species pairs, which share no evolutionary history, were chosen to ensure independence and thus a valid test of repeatability.

A nonparametric statistical test of repeatability was used to avoid assumptions about possible distributions of rates of molecular evolution. For both comparisons each window was ranked across the gene. Then the two patterns of ranks across the gene were compared using a rank correlation test. Thus a statistic was obtained (P_{rank}) to describe the probability of the ranking patterns being so similar through chance alone.

Analysis of similarity: For all of the nonsynonymous substitutions in the two comparisons, we characterized the extent to which these changes were conservative. This provides an indication as to whether it is likely that the replacements seen might be owing to weak stabilizing selection (in which case conservative changes are likely) or due to positive selection (in which case nonconservative changes are more likely; *cf.* TUCKER and LUNDRIGAN 1995). To characterize changes as being conservative or not we employed PAM matrices. Such matrices use data on the difference between the expected number of replacements of one amino acid with another (from knowledge of relative positions in the genetic code), and data on the actual rate at which such substitutions occur. The difference in the two rates is thought to indicate some measure of physicochemical similarity between the two amino acids under consideration such that similar amino acids are more likely to be interchangeable.

Different PAM matrices are appropriate for genes of different levels of divergence. We obtained appropriate matrices from the Amino Acid Index Database (http://www.cbi.pku.edu.cn/srs5bin/cgi-bin/wgetz?-fun+Pagelibinfo+-info+AAINDEX/). As the mouse-rat protein comparison shows 93% identity, we employed the BENS940101 PAM6-8 matrix (BENNER *et al.* 1994). The human-cow protein comparison shows 81% identity, so we used the BENS940102 PAM22-29 matrix (BENNER *et al.* 1994). For each segment in the gene each amino acid

difference was assigned the relevant value from the appropriate PAM matrix. The mean value per amino acid change per segment could then be calculated. We can then compare the regions of interest with the evolution in the rest of the gene.

Signal sequence evolution: To ask whether the signal sequence of Igf2r has an unusual mode of evolution we compared its evolution with that of comparable signal sequences in other genes. We used NCBI Entrez (http://www.ncbi.nlm.nih.gov/Entrez/) and the HOVERGEN database (Duret *et al.* 1994) to find nine other genes for which signal sequence information was available for mouse and rat orthologues. Alignments were prepared using the Wisconsin package (Genetics Computer Group 1994). DNA alignments were prepared for both signal sequences and entire protein-coding sequences using the default settings of PILEUP. The program GAPFRAME was used to move all gaps to codon boundaries. Nucleotide substitution rates were estimated with the Moriyama package (Moriyama and Powell 1997), using the synonymous/nonsynonymous method of Li (1993) and the multiple substitution correction method of Tamura and Nei (1993).

The genes (accession numbers in brackets, mouse then rat) were as follows: interleukin 4 receptor (M29854, X69903), low density lipoprotein receptor (X64414, X13722), tumor necrosis factor receptor 1 (M60468, M63122), calreticulin (X14926, X53363), cholesterol esterase (U33169, X16054), glutamate dehydrogenase (X57024, X14223), procathepsin E (X97399, D38104), immunoglobulin light chain V-λ-1 (J00590, M17092), and oxytocin-neurophysin I (M88355, K01701).

RESULTS

Repeatability of K_a/K_s: With the 24 nonoverlapping sections, K_a/K_s patterns across *Igf2r* are significantly repeatable ($P_{rank} < 0.01$, see Figure 1). It appears that the large-scale fluctuations in K_a/K_s rather than the small-scale fluctuations are responsible for the observed repeatability (our unpublished results). K_a/K_s is an indicator of both the intensity and type of selection. Stronger stabilizing selection reduces K_a/K_s, while both weaker stabilizing selection and stronger positive selection increase K_a/K_s. A K_a/K_s ratio significantly above one is the strict requirement for evidence of positive selection, but K_a/K_s ratios less than one do not preclude positive selection because different sites within the same segment might be under different selection pressures. The repeatability of patterns of selection across *Igf2r* means that selection is able to respond differently to various small subdomains across the *Igf2r* protein. This result is consistent with the conservation of the various functions of the different parts of the Igf2r molecule across the cow, human, mouse, and rat species.

The significant repeatability of K_a/K_s across the *Igf2r* gene means that the unexpected K_a/K_s patterns are not artifacts, but are representative of *Igf2r*'s molecular evolution. We therefore need to explain the low rate of protein evolution at the Igf2 binding site, and the high rates of protein evolution both at the signal sequence and at one of the M6P binding sites.

Similarity of amino acid substitutions: To examine the selection pressures acting on the fast-evolving signal sequence and M6P binding site, we compared the amino

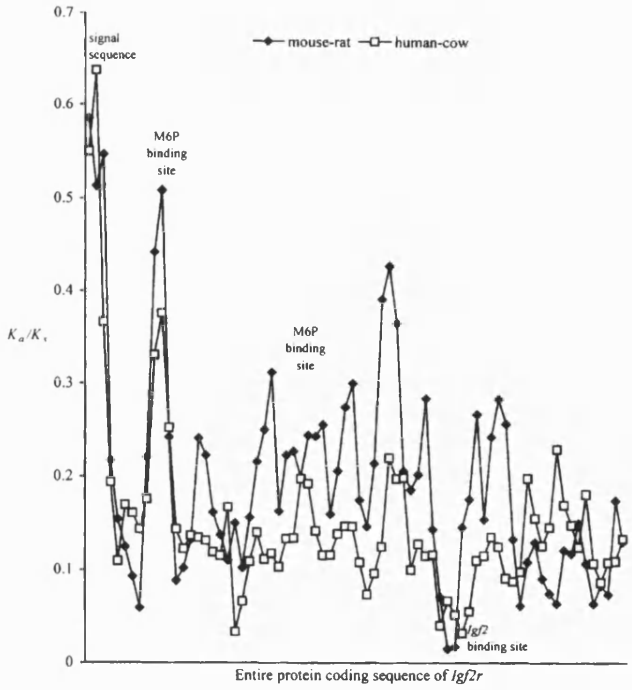


Figure 1.—K_a/K_s patterns are repeatable across *Igf2r*. The nonsynonymous over synonymous substitution rates (K_a/K_s) across the *Igf2r* gene are shown for both the mouse-rat and human-cow comparisons. The data are for the 74 overlapping windows. Note the repeated peaks at the signal sequence and the first M6P binding site, and the repeated troughs at the Igf2 binding site.

acids that had changed in each nonoverlapping window to ask how conservative the changes were, using data from the appropriate PAM matrices to ascribe similarity.

In the human-cow comparison the changes in the signal sequence were on average exceptionally nonconservative (see Figure 2). Indeed, none of the windows shows a lower level of similarity and the average level

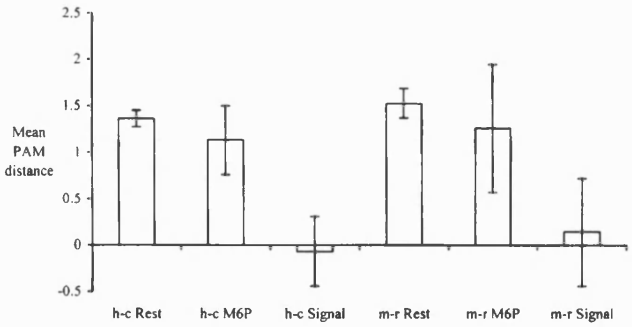


Figure 2.—Mean PAM distances per amino acid change for different regions of *Igf2r*. Data are presented for both the human-cow (h-c) and mouse-rat (m-r) comparisons. The human-cow and mouse-rat data were obtained using the PAM22-29 and PAM6-8 matrices, respectively (see materials and methods for references). Signal, the window containing the signal sequence; M6P, the two windows containing the rapidly evolving mannose-6-phosphate binding site; and Rest, all other parts of the gene. Bars show SE.

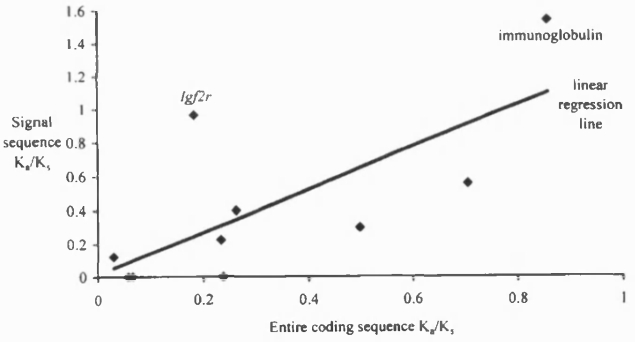


FIGURE 3.—K_a/K_s of signal sequences *vs.* K_a/K_s of entire genes. A comparison is shown of the mouse-rat K_a/K_s values of the signal sequence and of the entire protein coding sequence for ten genes (full gene names and accession numbers in MATERIALS AND METHODS). Both *Igf2r* and the immunoglobin gene show higher signal sequence K_a/K_s values than the other genes.

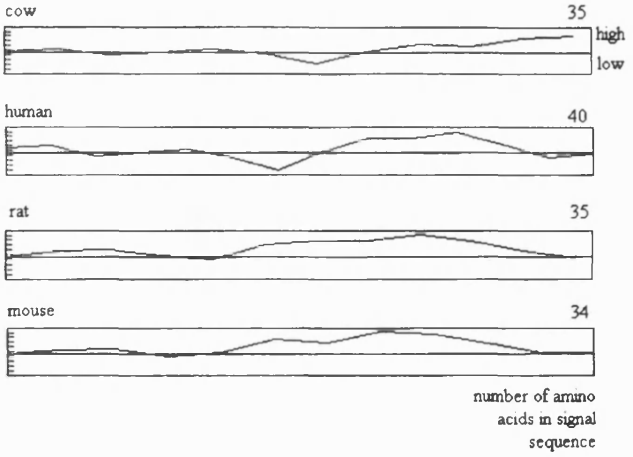


FIGURE 4.—Hydrophobicity plots across *Igf2r* for four species. Sliding window plots of hydrophobicity values [as produced by STADEN's (1996) package] across the signal sequences of the *Igf2r* gene in mouse, rat, cow, and human are shown. The signal sequence of cow is the only one that does not show the prototypical pattern of a central peak. The numbers give the lengths of the different signal sequences in amino acids (showing that the plots are not aligned).

of similarity in the gene as a whole is significantly higher ($P = 0.0017$, two-tailed Mann-Whitney *U*-test). This suggests that positive selection might be acting on this sequence. In the mouse-rat comparison the picture is not so clear. The signal sequence does show a low similarity but three other windows have lower values. Compared with the rest of the gene the signal sequence does have a lower value ($P = 0.0387$, two-tailed Mann-Whitney *U*-test).

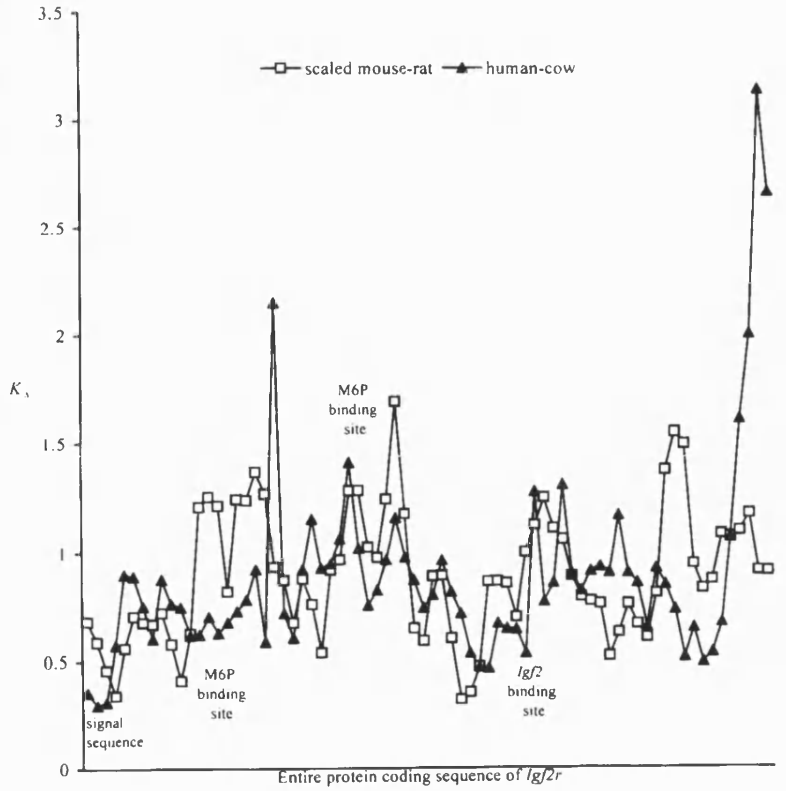
At the M6P binding site there appears to be nothing unusual about the amino acids being replaced in both mouse-rat and human-cow, suggesting that this region is under weak stabilizing selection (two-tailed Mann-Whitney *U*-tests between the M6P site and the rest of the gene give $P = 0.6248$ and $P = 0.6829$ for the human-cow and mouse-rat comparisons, respectively).

The signal sequence of Igf2r is unusually fast evolving: One would expect signal sequences to be generally slow evolving (see Introduction), but is this actually true; *i.e.*, Is the signal sequence of *Igf2r* unusually fast evolving? When signal sequence K_a/K_s is plotted against the entire coding sequence K_a/K_s for the nine genes and for *Igf2r* it becomes clear that *Igf2r*'s signal sequence is evolving unusually fast (see Figure 3). The only gene with a comparable signal sequence rate is immunoglobulin light chain V-λ-1. However, the rate of signal sequence evolution in this gene is comparable to the rate of evolution in the rest of the gene. *Igf2r* is hence unusual in that the signal sequence shows an unusually high ratio of K_a/K_s given the rate of evolution of the rest of the sequence. Compared with a large sample of genes (unpublished data) the K_a/K_s ratio of the majority of *Igf2r* is not unusual. The value in the signal sequence is high in these terms. Hence the discrepancy between the signal sequence and the rest of the gene is better understood as unusually fast evolution of the signal sequence rather than unusually slow evolution of the majority of the gene.

Signal sequence hydrophobicity: Is the rapid evolution of the *Igf2r* signal sequence at the expense of its functionality? We examined the hydrophobicity plots of the four species' signal sequences using the Staden package (STADEN 1996). The conserved signal sequence structure (see Introduction) predicts a central hydrophobic peak. The signal sequences of human, rat, and mouse fulfilled this prediction but the signal sequence of cow showed no such feature (Figure 4). This observation could explain the high divergence in the human-cow comparison, but offers no clues as to why both the human-cow and mouse-rat signal sequence divergences are so high.

Repeatability of K_s: With the 24 nonoverlapping sections, K_s variation is significantly repeatable ($P_{rank} = 0.05$; see Figure 5). With the rank of size change between neighboring overlapping sections as a potentially more powerful test, K_s patterns remained significantly repeatable ($P_{rank} < 0.05$). The repeatability of K_s across the *Igf2r* gene thus implies that the variation in K_s within the gene is being shaped by deterministic forces. We have performed several tests to attempt to distinguish between the potential deterministic forces described above (see Introduction). First, however, we ask whether the repeatability might be an artifact.

K_s repeatability is probably not a methodological artifact: Unless multiple substitution correction methods take account of variation in sequence composition, nucleotide substitution rate estimates can be biased (PESOLE *et al.* 1995). Furthermore, the use of different substitution rate estimation methods can lead to different conclusions using the same dataset (SMITH and HURST 1998). Could this effect be responsible for our observation of K_s repeatability, since KIMURA's (1980) two parameter



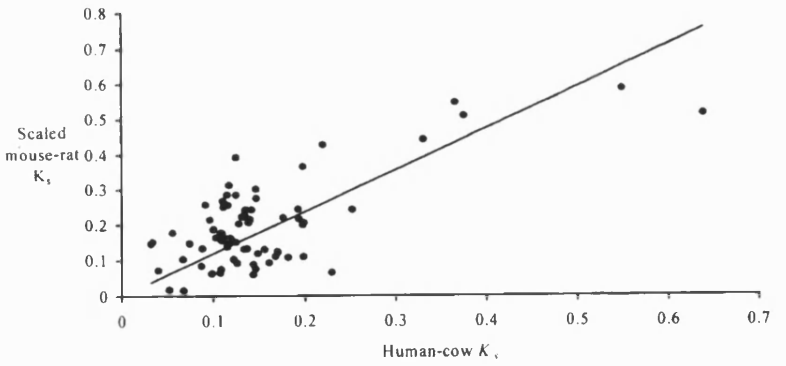


FIGURE 5.—K_s patterns are repeatable across *Igf2r*. The mouse-rat and human-cow synonymous substitution rates (K_s) across the *IGF2R* gene are shown in the upper plot. The data are of the 74 overlapping windows. The actual human-cow values are used, but the mouse-rat values are scaled up so that the means of both sets of K_s values are the same, to allow easier comparison. The lower plot shows the correlation between the mouse-rat and human-cow orthologous overlapping window K_s values.

model does not take account of variation in sequence composition? Under this hypothesis our finding of K_s repeatability would be an artifact caused by compositional repeatability leading to repeatable bias in K_s estimates. To test this idea, we reanalyzed the section alignments with Moriyama's package (MORIYAMA and POWELL 1997) using the synonymous/nonsynonymous method of LI (1993) and the multiple substitution correction method of TAMURA and NEI (1993; that uses all four nucleotide frequencies, as well as a transversion substitution rate and two transition substitution rates). A significant K_s repeatability was observed ($P_{rank} < 0.05$), and thus the repeatability of K_s is probably not a methodological artifact.

K_s repeatability is not due to the majority of Igf2r consisting of repeated units: The large extracellular domain of IGF2R consists almost entirely of 15 conserved repeat units, between which there is roughly 20% amino acid identity in the human gene (MORGAN *et al.* 1987). If compositional features affect K_s values, one might expect to see each repeat giving similar K_s patterns. Using the Wisconsin package (GENETICS COMPUTER GROUP 1994), alignments of the repeats were prepared using the same alignment gaps as for the preparation of overlapping and nonoverlapping sections. For each repeat, and for both species pairs, three alignments were prepared: one of the 5′ three-fifths of the repeat (section 1), one of the middle three-fifths (section 2), and one of the 3′

three-fifths (section 3). Overlapping sections were used because of the small size of the repeats (~150 bp). K_s was then estimated for each alignment using Moriyama's package (MORIYAMA and POWELL 1997).

It was then determined for each repeat in both species comparisons which of the three sections in each repeat (first, second, or third) had the highest K_s. The null hypothesis that there is no link between repeats and mutation rates predicts that in any given repeat all three thirds of the sequence are equally likely to have the fastest rate of evolution. The alternative hypothesis that the repeat structure does influence K_s would suppose that repeatably one of the three sections will have the highest K_s. Thus for both species comparisons the null expectation was that of the 15 repeats, five sections should have the first third having the highest K_s, five should have the second with highest K_s, etc. The human-cow comparison gave (6, 5, 4) while the mouse-rat comparison gave (7, 3, 5). Neither result was significantly different from the null expectation (the chi-squared test gave $P > 0.5$ for both comparisons). Thus the null hypothesis of no link between repeats and K_s patterns cannot be rejected, and so we conclude that the repeatability of K_s patterns is probably not due to the repeatability of molecular evolution in repeats. We note, however, that this is a relatively weak test.

Having established that the repeatability in K_s is probably not an artifact, we now turn to the question of why this repeatability is found. We consider three hypotheses (see Introduction for details): either (1) the variation is due to differences in composition along the gene allied with some form of mutational bias, or (2) the variation is due to selection favoring important sites to have low mutation rates (selected mutation rate hypothesis), or (3) K_s variation is the result of varying selection on codon usage (common constraints hypothesis). Hypotheses 1 and 2 assume that silent site mutations are neutral. Hypotheses 2 and 3 propose that the repeatability is selection-driven rather than mutation-driven. We shall start with these.

Adaptive explanations for K_s repeatability: Here we evaluate the extent to which adaptive explanations for K_s variability and repeatability can explain the data. The common constraints argument predicts that codon bias should be negatively correlated with K_s [because both depend on the strength of selection, for Drosophila and bacterial data see LI (1997)], and that there should be a correlation between K_a and K_s (because both are influenced by similar selective pressures). The selected mutation rate argument predicts that regions of the gene under strong stabilizing selection (low K_a) should have low mutation rates (low K_s), and conversely regions under weak stabilizing selection or even positive selection (high K_a) should have high mutation rates (high K_s). For both adaptive hypotheses, K_s repeatability follows from selection patterns being conserved across all four species. Thus the selected mutation rate argument

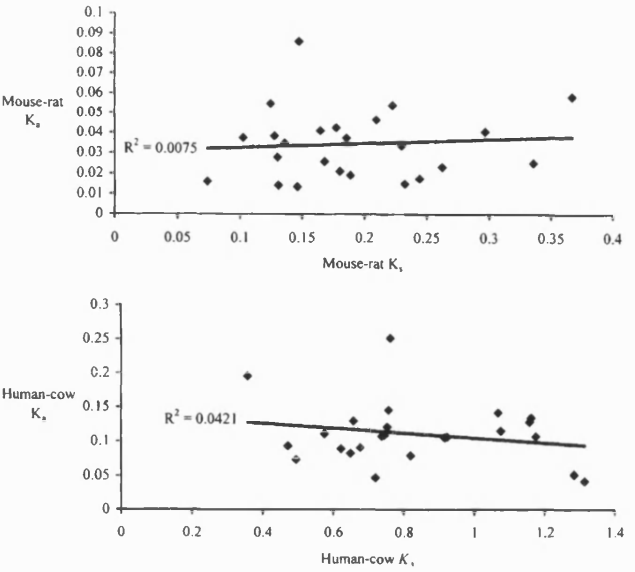


FIGURE 6.—No intragenic correlation between K_a and K_s. The K_a value of each nonoverlapping section is plotted against the corresponding K_s value for both the mouse-rat and the human-cow comparisons. The linear regression lines for both comparisons indicate no significant positive correlation between K_a and K_s.

agrees with the common constraints argument in predicting K_s repeatability and a correlation between K_a and K_s, but differs in predicting no negative correlation between codon bias and K_s.

Evidence from codon bias data: To differentiate between the selected mutation rate and common constraints hypotheses, we looked for a relationship between codon bias and K_s within the *Igf2r* gene. We used the Moriyama package (MORIYAMA and POWELL 1997) to determine effective numbers of codons (*ENC*; WRIGHT 1990) for the sections. For both species comparisons the *ENC* data of the two species were averaged. There was no correlation between mouse-rat *ENC* and K_s ($P_{rank} \gg 0.1$), but the human-cow comparison showed a negative correlation on the edge of significance ($P_{rank} = 0.05$). A negative correlation between *ENC* and K_s is equivalent to a positive correlation between codon bias and K_s (the lower the effective number of codons the stronger the codon bias), and so both species comparisons rejected the prediction of the common constraints hypothesis. The codon bias data show us that silent sites do not appear to be affected by selection, and thus K_s can be taken as an indicator of mutation rate.

No correlation between K_s and K_a: Both selective explanations of K_s repeatability predict a positive correlation between K_s and K_a. However, neither the mouse-rat nor the human-cow comparison give a significant positive correlation ($P_{rank} > 0.5$ for both; see Figure 6). This result provides evidence against both adaptive explanations of K_s repeatability. We note additionally, that presence of a significant positive correlation between K_a and

K_s could not have been taken as conclusive evidence in favor of the selected mutation rate argument because there are several other possible explanations for such a correlation (see Li 1997).

No correlation between K_s and functional importance: We separated the gene into sections according to predicted functional importance, and then examined whether regions likely *a priori* to be important (such as binding sites) showed a systematically lower K_s than the rest of the gene. Both the common constraints and selected mutation rate hypotheses would predict lower K_s values for more important sections of the gene because in such regions selection is likely to be more powerful. However, both the mouse-rat and human-cow comparisons gave a higher mean K_a/K_s for the extracellular ligand-binding regions than the extracellular regions not associated with ligand binding (0.27 *vs.* 0.17 and 0.20 *vs.* 0.11, respectively). This result contradicts the prediction of greater functional importance of binding sites (with stabilizing selection, the stronger the selection the lower the K_a/K_s values), and thus renders this test inapplicable. However, the observation that the K_s at the Igf2 binding site is not low despite the K_a/K_s being unusually low (see Figures 1 and 5) provides further, albeit circumstantial, evidence against both the common constraints and selected mutation rate hypotheses.

Mutational explanations for K_s repeatability: Here we consider two possible compositional correlates to K_s variation. First we ask whether GC composition might explain the patterns. Second we ask more particularly whether methylation at CpG sites might explain some of the variance.

GC composition: If local base composition influences mutation rates (and hence K_s), then the repeatability of K_s between highly related (and hence compositionally similar) sequences follows. For example, a correlation between GC content and K_s has been proposed (Wolfe *et al.* 1989). However, we found no significant correlation between fourfold site GC content and K_s for either the mouse-rat ($P_{rank} > 0.1$) or human-cow ($P_{rank} > 0.1$) comparisons. Wolfe *et al.* (1989) is now discredited due to (1) a small sample size effect (Wolfe and Sharp 1993) and (2) a biased estimating protocol (Bernardi *et al.* 1997). Correction for both finds no correlation between chromosomal site (defined by GC content) and mutation rate given by K_s (Bernardi *et al.* 1997), as we have also found (Smith and Hurst 1998). The finding that the repeats within the *Igf2r* gene do not show similar K_s plots (see above) also suggests that composition might not explain the patterns of K_s variation.

The influence of methylation: Even though local GC content seems not to influence local K_s, more complicated compositional characters may well influence local K_s (for example, see Morton *et al.* 1997), and thus the above result does not falsify the compositional hypothesis. One compositional feature that might well influence K_s patterns, and be responsible for the K_s repeatability under the compositional hypothesis, is the methylation pattern. Methylation is known to affect mutation rates with a methylated cytosine in a CpG pair 10 to 20 times more mutable than an unmodified cytosine (Kendrew 1994).

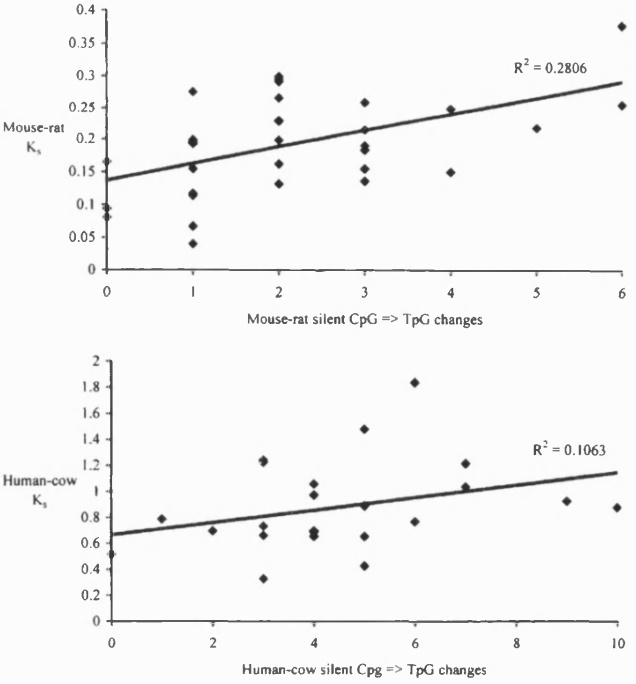


Figure 7.—Silent CpG ⇒ TpG changes correlate with K_s. For both comparisons of mouse-rat and human-cow, and for each nonoverlapping window, K_s is plotted against the number of silent CpG ⇒ TpG changes. The correlations show that such methylation-induced mutations provide a possible explanation for repeatable K_s patterns.

To test this possibility we examined the frequency of CpG ⇒ TpG mutations. For each segment we found the number of times a CpG was found in one of the two sequences and TpG was found in the other, where the C existed at a third site. This protocol gives all possible synonymous CpG ⇒ TpG changes (we assume all mutational changes to be in this direction). We find that the variation in K_s between segments is correlated to the frequency of such mutations (see Figure 7). The mouse-rat comparison shows a significant correlation ($P_{rank} = 0.02$), while the human-cow comparison shows a tendency in the same direction ($0.1 > P_{rank} > 0.05$). We also find that this class of mutations accounts for a reasonable proportion of all mutations (15% in the human-cow comparison, and 21% in the mouse-rat comparison). We conclude that methyl-induced mutations can account for some of the repeatability that we observe.

Two hypotheses might be considered to account for this pattern. At one extreme, all segments might have an equal number of CpG sites that might change, but these segments differ in the density of methylation. Alternatively, the segments might have the same density of methylation per CpG, but differ in the total number of CpG sites.

We find that variation in total number of mutable CpGs partially explains the pattern in K_s. We calculated the number of times CpG was found in both sequences (*i.e.*, the CpG/CpG frequency) again with the C at the third site. Adding the CpG/CpG figure to the CpG/TpG figure we arrive at a figure for the total number of mutable or mutated silent CpG sites. This total correlates well with K_s ($P_{rank} < 0.05$ for the mouse-rat comparison, though $P_{rank} > 0.1$ for the human-cow comparison), indicating that the variance in K_s explained by methylation is to some extent a result of differences between segments in CpG content.

Conversely, we can ask whether segments also differ in their propensity to methylate CpG sites. The methylation of *Hpa*II sites of the mouse E15 embryo *Igf2r* gene has been shown to vary from 50% to 90% throughout the locus (Stoger *et al.* 1993), which at least suggests that there might be enough variation in methylation patterns to account for the highly variable mutation rate patterns (Figure 5). Asking whether segments differ significantly in the proportion of CpGs that become TpGs we find heterogeneity between segments (chi-squared test for heterogeneity gives $P < 0.05$ for the mouse-rat comparison, and $0.5 > P > 0.1$ for the human-cow comparison). This supports the view that the observed variation in CpG > TpG changes (and hence K_s) is at least partially due to differences in the methylation density of potentially methylated sites.

This finding contrasts with human polymorphism data. The M6P/IGF2R Information Core (http://www.radonc.duke.edu/~jirtle/homepage.html) provides a human mutation database for *IGF2R*. Six substitutional mutations resulting in disease symptoms are described, with no C to T mutations. However, this small dataset suffers from an ascertainment bias, in that not all substitutional mutations are equally likely to cause disease symptoms.

DISCUSSION

The mammalian *Igf2r* gene is multifunctional, and plays an important role in glycoprotein transport via M6P binding (Kornfeld 1992) and growth and development via IGF2 and RA binding (Kang *et al.* 1998), and also tumor suppression (Desouza *et al.* 1997). In keeping with this functional complexity mammalian *Igf2r* displays complex patterns of molecular evolution. The significant repeatability of both K_a/K_s and K_s means that these patterns are driven by deterministic forces, and we have tested a number of hypotheses concerning the nature of these forces.

K_a/K_s variation and the conflict theory of imprinting: A simple prediction of the conflict theory of imprinting is that the Igf2 binding site should exhibit a high rate of molecular evolution driven by positive selection (McVean and Hurst 1997b; Trivers and Burt 1998). Instead we find evidence from K_a/K_s patterns for strong stabilizing selection at the Igf2 binding site in both the mouse-rat and human-cow comparisons. This finding cannot then be dismissed as a statistical artifact. Fast evolution at this site might have provided good evidence supporting the conflict hypothesis (McVean and Hurst 1997b; Trivers and Burt 1998). We have looked for positive selection in other parts of the gene, and have found evidence from PAM matrix data that the signal sequence might be under positive selection. We have shown that signal sequence evolution in *Igf2r* is exceptional when compared to other examples of signal sequence evolution. This is the first evidence for any form of molecular evolution (*sensu stricto*) in an imprinted gene that is different from the form of molecular evolution in most nonimmune system/nonplacental genes [note, however, that imprinted genes have a few other unusual properties, such as intronic dimensions (Hurst *et al.* 1996), sex specific recombination rates (Paldi *et al.* 1995), and repeat structures (Neumann *et al.* 1995)].

The evolution of *Igf2r*'s signal sequence might be a result of antagonistic coevolution, in which case we can speculate that the signal sequence might be coevolving with the gene product of an imprinted gene (or genes) expressed from the paternally derived genome. The antagonism might concern the location of Igf2r. The putative paternally derived product could, in theory, attempt to remove the protein to a cellular location where its effects on Igf2 levels are minimal. It would be interesting to see the effect of site-directed mutagenesis of the signal sequence on the cellular location of Igf2r (no mutations in the *IGF2R* signal sequence causing human disease are listed at the M6P/IGF2R Information Core at http://www.radonc.duke.edu/~jirtle/homepage.html). Alternatively, the unusual evolution at this site might be a property of *Igf2r* independent of imprinting. It would then be informative to analyze the molecular evolution of *Igf2r* in birds for example, assuming that avian *Igf2r* is not imprinted (as yet, chicken is the only bird for which the *Igf2r* gene has been sequenced).

Both this study and a previous analysis of the K_a/K_s values of imprinted genes (McVean and Hurst 1997b) assume that the conflict theory of imprinting predicts rapid protein evolution at the sites of functional antagonism. This assumption is in keeping with the results of studies of classic parent-offspring conflict (see Introduction), though such fast rates of evolution might be due to the acquisition of new roles. However, Haig (1997) has suggested an alternative prediction, that the mutual dependence of the roles of Igf2r and Igf2 leads to an ESS rather than a continuing arms race. Haig suggests that the *a priori* expectation is unclear. This failure to provide a testable prediction of the rates of evolution at sites of functional antagonism might be considered a weakness of the current models of the conflict theory of imprinting.

K_s variation and the evolution of mutation rates: The result of K_s repeatability in mammalian *Igf2r*, along with

evidence from codon bias data that the silent sites appear to be neutral, extends our understanding of mutation rate variation: to quote CASANE *et al.* (1997) "it is still not clear whether the mutation pattern is constant over the entire genome of an organism or is variable among regions of the genome." Large scale variation in mammalian mutational patterns has been previously demonstrated on two levels: between different chromosomes [K_s used as a measure of mutation rate in MCVEAN and HURST (1997a)] and between pseudogenes inserted into different genomic regions (CASANE *et al.* 1997). The repeatability of K_s patterns within *Igf2r* provides strong evidence that there is deterministic variation in mutation rates at a fine scale, between different regions within the same gene.

We have tested various hypotheses to explain the K_s repeatability. Since the K_s repeatability does not appear to be a methodological artifact, the remaining hypotheses can be divided into two groups, mutational and adaptive explanations. The selected mutation rate argument holds that local mutation rates, given by K_s, are the results of conflicting selection pressures governing mutation rates (see Introduction). Alternatively, the common constraints argument proposes that silent sites may be under selection, and so K_s is not the mutation rate but instead depends on selective constraints (see Introduction). The lack of a negative correlation between codon bias and K_s provides evidence against the common constraints hypothesis, and the lack of a correlation between K_s and K_a provides evidence against both adaptive explanations. The comparison of local mutation rates and local functional importance provided further, but circumstantial, evidence against both adaptive hypotheses.

Repeatable composition of repeats does not cause the repeatable K_s patterns. We did, however, find that silent C $\Rightarrow$ T mutations at CpG sites do covary with K_s, suggesting that the variation in K_s is somehow connected with methylation-induced mutational patterns. The variation in silent C $\Rightarrow$ T mutations at CpG sites appears to be the result of variation both in the number of mutable sites, and in the density of methylation.

It has previously been reported that, at least for GC-rich genes, the variation in K_s does covary with the variation in K_a (ALVAREZ-VALIN *et al.* 1998). This was interpreted as evidence for selection differentially affecting codon usage across the gene. Were this the only force affecting K_s variation, then one would not have predicted K_s repeatability in a gene, such as *Igf2r*, in which K_a and K_s do not covary. We hence conclude that there must be some other forces determining variation in K_s. Variation in silent CpG mutation appears in our case to be an important variable and it is noteworthy that analysis of GC content alone did not detect this.

We thank Etsuko Moriyama and Josep Comeron for providing us with their programs for evolutionary analysis, Gil McVean for suggestions concerning methodology, David Haig for discussions on imprinting, Denise Barlow for advice on methylation data, and two anonymous referees for suggesting many improvements.

LITERATURE CITED

ALTSCHUL, S. F., W. GISH, W. MILLER, E. W. MYERS and D. J. LIPMAN, 1990 Basic local alignment search tool. J. Mol. Biol. **215:** 403–410.

ALVAREZ-VALIN, F., K. JABBARI and G. BERNARDI, 1998 Synonymous and nonsynonymous substitutions in mammalian genes: intragenic correlations. J. Mol. Evol. **46:** 37–44.

BENNER, S. A., M. A. COHEN and G. H. GONNET, 1994 Amino-acid substitution during functionally constrained divergent evolution of protein sequences. Prot. Eng. **7:** 1323–1332.

BERNARDI, G., D. MOUCHIROUD and C. GAUTIER, 1997 Isochores and synonymous substitutions in mammalian genes, pp. 137–168 in *DNA and Protein Sequence Analysis*, edited by M. J. BISHOP and C. J. RAWLINGS. Oxford University Press, Oxford.

CASANE, D., S. BOISSINOT, B. H. J. CHANG, L. C. SHIMMIN and W. H. LI, 1997 Mutation pattern variation among regions of the primate genome. J. Mol. Evol. **45:** 216–226.

CHORNEY, M. J., K. CHORNEY, N. SEESE, M. J. OWEN, J. DANIELS *et al.*, 1998 A quantitative trait locus associated with cognitive ability in children. Psychol. Sci. **9:** 159–166.

COMERON, J. M., 1995 A method for estimating the numbers of synonymous and nonsynonymous substitutions per site. J. Mol. Evol. **41:** 1152–1159.

DENNIS, P. A., and D. B. RIFKIN, 1991 Cellular activation of latent transforming growth factor beta requires binding to the cation-independent mannose 6-phosphate/insulin-like growth factor type II receptor. Proc. Natl. Acad. Sci. USA **88:** 580–584.

DESOUZA, A. T., T. YAMADA, J. J. MILLS and R. L. JIRTLE, 1997 Imprinted genes in liver carcinogenesis. FASEB J. **11:** 60–67.

DRAKE, J. W., 1991 A constant rate of spontaneous mutation in DNA-based microbes. Proc. Natl. Acad. Sci. USA **88:** 7160–7164.

DURET, L., D. MOUCHIROUD and M. GOUY, 1994 HOVERGEN—a database of homologous vertebrate genes. Nucleic Acids Res. **22:** 2360–2365.

EFSTRATIADIS, A., 1994 Parental imprinting of autosomal mammalian genes. Curr. Opin. Genet. Dev. **4:** 265–280.

GENETICS COMPUTER GROUP, 1994 Program Manual for the Wisconsin Package, Version 8, September 1994, Genetics Computer Group, 575 Science Drive, Madison, WI.

HAIG, D., 1997 Parental antagonism, relatedness asymmetries, and genomic imprinting. Proc. R. Soc. Lond. B **264:** 1657–1662.

HUGHES, A. L., M. K. HUGHES, C. Y. HOWELL and M. NEI, 1994 Natural-selection at the class-II major histocompatibility complex loci of mammals. Philos. Trans. R. Soc. Lond. Ser. B Biol. Sci. **346:** 359–366.

HURST, L. D., 1997 Evolutionary theories of genomic imprinting, pp. 211–237 in *Frontiers in Molecular Biology: Genomic Imprinting in Mammals*, edited by W. REIK and A. SURANI. Oxford University Press, Oxford.

HURST, L. D., and G. T. MCVEAN, 1997 Growth effects of uniparental disomies and the conflict theory of genomic imprinting. Trends Genet. **13:** 436–443.

HURST, L. D., G. T. MCVEAN and T. MOORE, 1996 Imprinted genes have few and small introns. Nat. Genet. **12:** 234–237.

IZARD, J. W., S. L. RUSCH and D. A. KENDALL, 1996 The amino-terminal charge and core region hydrophobicity interdependently contribute to the function of signal sequences. J. Biol. Chem. **271:** 21579–21582.

KANG, J. K., Y. LI and A. LEAF, 1998 Mannose-6-phosphate/insulin-like growth factor-II receptor is a receptor for retinoic acid. Proc. Natl. Acad. Sci. USA **95:** 13671–13676.

KENDREW, J. (Editor), 1994 *The Encyclopedia of Molecular Biology.* Blackwell, Oxford.

KIMURA, M., 1980 A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. J. Mol. Evol. **16:** 111–120.

KIMURA, M., 1983 *The Neutral Theory of Evolution.* Cambridge University Press, Cambridge, UK.

KORNFELD, S., 1992 Structure and function of the mannose 6-phos-

phate insulin-like growth factor-II receptors. Annu. Rev. Biochem. **61**: 307–330.

LI, W. H., 1993 Unbiased estimation of the rates of synonymous and nonsynonymous substitution. J. Mol. Evol. **36**: 96–99.

LI, W. H., 1997 *Molecular Evolution*. Sinauer Associates, Sunderland, MA.

MCVEAN, G. T., and L. D. HURST, 1997a Evidence for a selectively favourable reduction in the mutation rate of the X chromosome. Nature **386**: 388–392.

MCVEAN, G. T., and L. D. HURST, 1997b Molecular evolution of imprinted genes: no evidence for antagonistic coevolution. Proc. R. Soc. Lond. Ser. B Biol. Sci. **264**: 739–746.

MILLS, J. J., J. G. FALLS, A. T. DESOUZA and R. C. JIRTLE, 1998 Imprinted MGP/Igf2 receptor is mutated in rat liver tumours. Oncogene **16**: 2797–2802.

MOCHIZUKI, A., Y. TAKEDA and Y. IWASA, 1996 The evolution of genomic imprinting. Genetics **144**: 1283–1295.

MOORE, T., and D. HAIG, 1991 Genomic imprinting in mammalian development: a parental tug-of-war. Trends Genet. **7**: 45–49.

MORGAN, D. O., J. C. EDMAN, D. N. STANDRING, V. A. FRIED, M. C. SMITH *et al.*, 1987 Insulin-like growth factor-ii receptor as a multifunctional binding-protein. Nature **329**: 301–307.

MORIYAMA, E. N., and J. R. POWELL, 1997 Synonymous substitution rates in *Drosophila*: Mitochondrial versus nuclear genes. J. Mol. Evol. **45**: 378–391.

MORTON, B. R., V. M. OBERHOLZER and M. T. CLEGG, 1997 The influence of specific neighboring bases on substitution bias in noncoding regions of the plant chloroplast genome. J. Mol. Evol. **45**: 227–231.

MOUCHIROUD, D., C. GAUTIER and G. BERNARDI, 1995 Frequencies of synonymous substitutions in mammals are gene-specific and correlated with frequencies of nonsynonymous substitutions. J. Mol. Evol. **40**: 107–113.

NEUMANN, B., P. KUBICKA and D. P. BARLOW, 1995 Characteristics of imprinted genes. Nat. Genet. **9**: 12–13.

OGAWA, O., D. M. BECROFT, I. M. MORISON, M. R. ECCLES, J. E. SKEEN *et al.*, 1993 Human insulin-like growth-factor type-I and type-II receptors are not imprinted. Hum. Mol. Genet. **2**: 2163–2165.

PALDI, A., G. GYAPAY and J. JAMI, 1995 Imprinted chromosomal regions of the human genome display sex-specific meiotic recombination frequencies. Curr. Biol. **5**: 1030–1035.

PESOLE, G., G. DELLISANTI, G. PREPARATA and C. SACCONE, 1995 The importance of base composition in the correct assessment of genetic-distance. J. Mol. Evol. **41**: 1124–1127.

SMITH, N. G. C., and L. D. HURST, 1998 Sensitivity of patterns of molecular evolution to alterations in methodology. J. Mol. Evol. (in press).

SPENCER, H. G., M. W. FELDMAN and A. G. CLARK, 1998 Genetic conflicts, multiple paternity and the evolution of genomic imprinting. Genetics **148**: 893–904.

STADEN, R., 1996 The Staden sequence-analysis package. Mol. Biotech. **5**: 233–241.

STOGER, R., P. KUBICKA, C. G. LIU, T. KAFRI, A. RAZIN *et al.*, 1993 Maternal-specific methylation of the imprinted mouse *Igf2r* locus identifies the expressed locus as carrying the imprinting signal. Cell **73**: 61–71.

SUTTON, K. A., and M. F. WILKINSON, 1997 Rapid evolution of a homeodomain: evidence for positive selection. J. Mol. Evol. **45**: 579–588.

TAMURA, K., and M. NEI, 1993 Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. Mol. Biol. Evol. **10**: 512–526.

TRIVERS, R., and A. BURT, 1998 Kinship and genomic imprinting, in *Genomic Imprinting*, edited by R. OHLSSON. Springer, Heidelberg (in press).

TRIVERS, R. L., 1974 Parent-offspring conflict. Am. Zool. **14**: 249–264.

TSAUR, S. C., and C. I. WU, 1997 Positive selection and the molecular evolution of a gene of male reproduction, Acp26Aa of *Drosophila*. Mol. Biol. Evol. **14**: 544–549.

TUCKER, P. K., and B. L. LUNDRIGAN, 1995 The nature of gene evolution on the mammalian Y-chromosome—lessons from SRY. Philos. Trans. R. Soc. Lond. Ser. B Biol. Sci. **350**: 221–227.

WALLIS, M., 1981 The molecular evolution of pituitary growth hormone prolactin and plactental lactogen: a protein family showing variable rates of evolution. J. Mol. Evol. **17**: 10–18.

WALLIS, M., 1993 Remarkably high rate of molecular evolution of ruminant placental lactogens. J. Mol. Evol. **37**: 86–88.

WALLIS, M., 1994 Variable evolutionary rates in the molecular evolution of mammalian growth hormones. J. Mol. Evol. **38**: 619–627.

WOLFE, K. H., and P. M. SHARP, 1993 Mammalian gene evolution—nucleotide-sequence divergence between mouse and rat. J. Mol. Evol. **37**: 441–456.

WOLFE, K. H., P. M. SHARP and W. H. LI, 1989 Mutation-rates differ among regions of the mammalian genome. Nature **337**: 283–285.

WRIGHT, F., 1990 The effective number of codons used in a gene. Gene **87**: 23–29.

XIE, S., J. GREEN, J. BIXBY, B. SZAFRANSKA, J. DEMARTINI *et al.*, 1997 The diversity and evolutionary relationships of the pregnancy-associated glycoproteins, an aspartic proteinase subfamily consisting of many trophoblast-expressed genes. Proc. R. Soc. Lond. Ser. B Biol. Sci. **94**: 12809–12816.

XU, Y. Q., C. G. GOODYER, C. DEAL and C. POLYCHRONAKOS, 1993 Functional polymorphism in the parental imprinting of the human IGF2R gene. Biochem. Biophys. Res. Commun. **197**: 747–754.

ZHOU, M., Z. M. MA and W. S. SLY, 1995 Cloning and expression of the cDNA of chicken cation-independent mannose-6-phosphate receptor. Proc. Natl. Acad. Sci. USA **92**: 9762–9766.

Communicating editor: G. B. GOLDING

Research Paper 5. Molecular evolutionary evidence that H19 mRNA is functional

Laurence Hurst and Nick Smith (1999)
Trends Genet. **15** 134-135.

Molecular evolutionary evidence that *H19* mRNA is functional

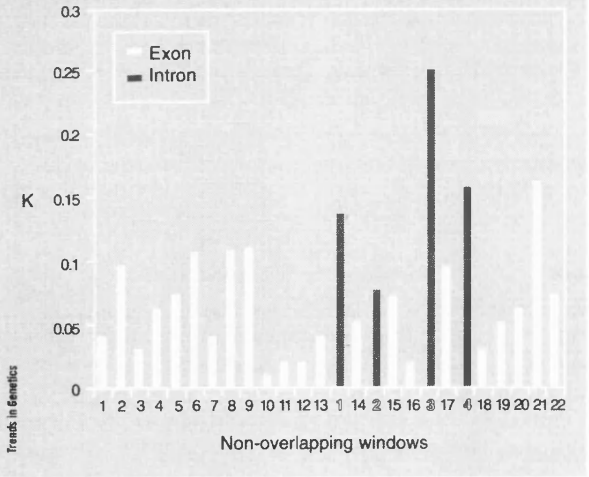
A few mammalian genes are known to have expression dependent upon the sex of the parent from which they were inherited[1]. While most of these imprinted genes described in human and mouse code for proteins, a few [*H19* (Ref. 2), *Xist* (Ref. 3) and *IPW* (Ref. 4)] code only for RNA. Here we investigate the molecular evolution of one of these, *H19*, so as to determine whether the spliced RNA might have function.

The prevailing view[5] is that the cytoplasmic *H19* transcript does not have any function, although the situation is by no means resolved. The murine knockout has a growth phenotype[6] but it is uncertain whether this is related to the activity of the transcript, given that over-expression of the gene has no growth phenotype[7]. Furthermore, while it was speculated that the RNA may function in regulating *Igf2* imprinting[2], the replacement of the *H19* structural gene with a protein-coding gene shows this not to be so[8]. Similarly, a tumour suppressing role has been claimed[9], but these results could not be reproduced[10]. Most recently[5], it has been shown that the *H19* transcript is associated with polysomes and may be an antagonist of *Igf2* expressivity *in trans*. Here we show that molecular evolutionary comparison of the mouse and rat versions of the gene indicate that the RNA is under stabilizing selection and hence is most likely functional.

To show that a protein coding region is functional and under stabilizing selection, one would typically show that the rate at which the protein evolves is very much lower than the rate of evolution of the DNA at degenerate sites, that is, those sites where changes do not affect the protein. For noncoding RNAs this analysis is not possible. Instead we have performed two tests. First, we have compared the rate of evolution per base pair within the exons of *H19* with the rate of evolution at fourfold degenerate sites (i.e. those where all four nucleotides at the third site specify the same amino acid) in a large number of orthologous mouse–rat genes (orthology has been ensured through checking in Hovergen[11]). Assuming that the substitution rate at fourfold degenerate sites is a measure of the rate of evolution on exon sites that are free from selection, we can then ask whether the rate of evolution in *H19* is what one would expect were the sequence randomly selected from a sample of neutrally evolving genes. From a sample of 432 orthologous genes, we find that only three have a lower fourfold site substitution rate than *H19* has across all the bases in its exons ($P = 0.007$). Some of the variance in the distribution of fourfold site evolution is the result of differences in the sizes of the genes concerned (inclusion of small genes increases the variance). Restricting our analysis to the half of our data set with the longest complete coding sequence, we find now that no genes have lower rates of evolution ($P < 0.005$). The mean total exonic size in this large set is comparable to that in *H19* (although note that in coding sequences less than a third of all sites are fourfold degenerate ones, whereas we examine all sites in the exons of *H19*) (Fig. 1).

The above result may either be because *H19* is under stronger stabilizing selection than fourfold degenerate sites or has a low mutation rate. The latter may be quite reasonable given that, as a class, imprinted genes tend to have low synonymous substitution rates[12]. Indeed, 2 of 15 imprinted genes have lower fourfold degenerate site substitution rates than *H19* (*Igf2* and *Neuronatin*), but both of these are very small sequences so the significance of this is unclear. More importantly, analysis of the evolution in the introns indicates that a low mutation rate across the gene does not alone explain why *H19* has such a very low substitution rate. We find that the introns of *H19* on the average evolve at 2.5 times the rate of the exons ($K_{exon} = 0.06$ per site, $K_{intron} = 0.15$) and that every intron has a higher substitution rate than both of its flanking exons (Fig. 1). Compared with a sample of 41 mouse–rat orthologous genes for which intronic data is available[13] only two have a higher ratio of intronic to exonic fourfold site rate ($P < 0.05$). Both of these have small total exonic content. This evidence, along with suggestive evidence of secondary structure conservation[14], is indicative of stabilizing selection acting on the exons. This in turn is consistent with *H19* mRNA having a function.

FIGURE 1. Substitution rates across *H19*



The exonic parts of the gene are split into non-overlapping windows 100 bp long (numbered 1 to 22). The introns are analysed whole and numbered in order. We used ClustalW[15] to align the intronic and exonic sequences of *H19* reported for rat (X59864) and mice (AF049091). The Y axis is the rate of substitution (K) using Tamura92[16] as the method of adjusting for multiple hits. Note that all introns have higher substitution rates than their flanking exons.

Acknowledgements

We thank two anonymous referees, A. Ward, M. Charalambous and G. Dell for their comments on earlier versions of the manuscript.

Laurence D. Hurst
l.d.hurst@bath.ac.uk

Nick G.C. Smith
bspns@bath.ac.uk

Department of Biology and Biochemistry, University of Bath, Claverton Down, Bath, UK BA2 4SD.

References

1 Morison, I.M. and Reeve, A.E. (1998) A catalogue of imprinted genes and parent-of-origin effects in humans and animals. *Hum. Mol. Genet.* 7, 1599–1609

2 Bartolomei, M.S. (1997) Function and epigenetic modification of the imprinted *H19* gene, in *Genomic Imprinting* (Reik, W. and Surani, A., eds), pp. 53–69, IRL Press

3 Kay, G.F. *et al.* (1994) Imprinting and X-chromosome counting mechanisms determine *Xist* expression in early mouse development. *Cell* 77, 639–650

4 Wevrick, R. *et al.* (1994) Identification of a novel paternally expressed gene in the Prader-Willi-Syndrome region. *Hum. Mol. Genet.* 3, 1877–1882

5 Li, Y.M. *et al.* (1998) The *H19* transcript is associated with polysomes and may regulate IGF2 expression *in trans*. *J. Biol. Chem.* 273, 28247–28252

6 Leighton, P.A. *et al.* (1995) Disruption of imprinting caused by deletion of the *H19* gene region in mice. *Nature* 375, 34–39

7 Pfeifer, K. *et al.* (1996) The structural *H19* gene is required for transgene imprinting. *Proc. Natl. Acad. Sci. U. S. A.* 93, 13876–13883

8 Jones, B.K. *et al.* (1998) *Igf2* imprinting does not require its own DNA methylation or *H19* RNA. *Genes Dev.* 12, 2200–2207

9 Hao, Y. *et al.* (1993) Tumor-suppressor activity of *H19* RNA. *Nature* 365, 764–767

10 Reid, L.H. *et al.* (1996) Localization of a tumor suppressor gene in 11p15.5 using the G401 Wilms' tumor assay. *Hum. Mol. Genet.* 5, 239–247

11 Duret, L. *et al.* (1994) Hovergen – a database of homologous vertebrate genes. *Nucleic Acids Res.* 22, 2360–2365

12 Smith, G.C. and Hurst, L.D. The causes of synonymous rate variation in the rodent genome: can substitution rates be used to estimate the sex bias in mutation rate? *Genetics* (in press)

13 Smith, N.G.C. and Hurst, L.D. (1998) Sensitivity of patterns of molecular evolution to alterations in methodology: a critique of Hughes and Yeager. *J. Mol. Evol.* 47, 493–500

14 Tilghman, S.M. *et al.* (1992) The mouse *H19* gene: its structure and function in mouse development, in *Nuclear Processes and Oncogenes* (Sharp, P.A., ed.), pp. 188–200, Academic Press

15 Thompson, J.D. *et al.* (1994) Clustal-W – improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* 22, 4673–4680

16 Tamura, K. (1992) Estimation of the number of nucleotide substitutions when there are strong transition-transversion and GC-content biases. *Mol. Biol. Evol.* 9, 678–687

An ethical collaboration

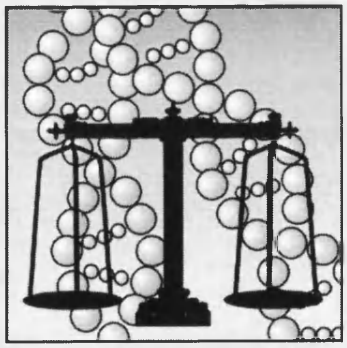
Ethics in Human Procreation, Genetic Diagnosis and Therapy, Sheffield, UK, 7–10 January 1999

The European Network for Biomedical Ethics convened earlier this year for the final symposium in a series of three, concerned with fundamental ethical questions and the social implications of human genetics and assisted reproduction. The symposium, organized by Deryck Beyleveld and Hille Haker, marked a collaboration between the Center for Ethics in the Sciences and Humanities, Tübingen, Germany (ZEW), and Sheffield Institute for Biotechnological Law and Ethics, UK (SIBLE). It brought together some 80 scientists and academics.

The symposium was opened at dinner on Thursday evening by Richard Cabourn MP (Sheffield Central, UK). He indicated that the British Government was considering setting up a committee to take direct action to review and revise the structure of advisory committees in the field of biotechnology. After dinner, Dietmar Mieth (ZEW), spoke, outlining the very real need for bringing social and ethical debate into the realm of science.

The symposium dealt with such varied subjects as procreation and parenthood; the moral protection of the human embryo and fetus; autonomy and recognition; the social implications of the new technologies; moral reasoning in applied ethics; and the legal regulations of assisted procreation, genetic diagnosis and therapy. In addition to the usual practice of inviting audience response to the papers, experts distinguished in their fields had been asked to prepare more detailed critiques of the main presentations, with a view to ensuring the highest quality of comment and criticism.

SIBLE



ZEW

All of the papers given were fascinating and insightful, but the contributions generating the most debate were those of Deryck Beyleveld (SIBLE); Marcus Düwell (ZEW); and the collaboration of Shaun Pattinson (SIBLE), and Deryck Beyleveld. Beyleveld, employing the rights theory of Alan Gewirth, attempted to clarify the nature and extent of rights had by the embryo/fetus, giving them as unassailable and concrete a basis as possible. Gewirth is unique among modern philosophers, although akin to Kant, in that he claims that right answers to moral questions are possible: rationality requires that a moral point of view be adopted, and rationality dictates the determinate content of moral behaviour, and the identity of moral beings. Beyleveld, although agreeing that the fundamental principle of Gewirth's theory is sound, has substantially modified the application of the theory. A claim to

Bev R. Clucas
LWP96BRC@
sheffield.ac.uk

Sheffield Institute of Biotechnological Law and Ethics, Sheffield University, Crookesmoor Building, Conduit Road, Sheffield, UK S10 1FL.

0168-9525/99/$ – see front matter PII: S0168-9525(99)01707-2

Research Paper 6. Can substitution rates be used to estimate the male to female mutation rate ratio in rodents?

Nick Smith and Laurence Hurst (1999)
Genetics **152** 661-673.

The Causes of Synonymous Rate Variation in the Rodent Genome: Can Substitution Rates Be Used to Estimate the Sex Bias in Mutation Rate?

Nick G. C. Smith and Laurence D. Hurst

Department of Biology and Biochemistry, University of Bath, Bath BA2 7AY, United Kingdom

Manuscript received September 24, 1998
Accepted for publication February 16, 1999

ABSTRACT

Miyata *et al.* have suggested that the male-to-female mutation rate ratio (α) can be estimated by comparing the neutral substitution rates of X-linked (X), Y-linked (Y), and autosomal (A) genes. Rodent silent site X/A comparisons provide very different estimates from X/Y comparisons. We examine three explanations for this discrepancy: (1) statistical biases and artifacts, (2) nonneutral evolution, and (3) differences in mutation rate per germline replication. By estimating errors and using a variety of methodologies, we tentatively reject explanation 1. Our analyses of patterns of codon usage, synonymous rates, and nonsynonymous rates suggest that silent sites in rodents are evolving neutrally, and we can therefore reject explanation 2. We find both base composition and methylation differences between the different sets of chromosomes, a result consistent with explanation 3, but these differences do not appear to explain the observed discrepancies in estimates of α. Our finding of significantly low synonymous substitution rates in genomically imprinted genes suggests a link between hemizygous expression and an adaptive reduction in the mutation rate, which is consistent with explanation 3. Therefore our results provide circumstantial evidence in favor of the hypothesis that the discrepancies in estimates of α are due to differences in the mutation rate per germline replication between different parts of the genome. This explanation violates a critical assumption of the method of Miyata *et al.*, and hence we suggest that estimates of α, obtained using this method, need to be treated with caution.

IT has long been thought that, at least in humans, most mutations originate in males, ever since HALDANE (1947) noted that most mutations causing Hemophilia A are paternally derived. The extent of any putative male mutation bias impacts on several areas including genetic counseling, understanding mutational processes, and many aspects of evolutionary biology (see HURST and ELLEGREN 1998). MIYATA *et al.* (1987) proposed to estimate the male-to-female mutation rate (α) by comparing nucleotide substitution rates in Y-linked, X-linked, and autosomal (A) genes. If the substitutions considered are selectively neutral (KIMURA 1983) and if multiple substitutions can be properly accounted for, then substitution rates can provide unbiased estimates of the mutation rate.

Previous comparisons of synonymous substitution rates on the X chromosome and the autosomes of mouse and rat have yielded estimates of rodent $\alpha = \infty$ (WOLFE and SHARP 1993; MCVEAN and HURST 1997). In contrast, comparisons of substitution rates on the X and Y chromosomes have provided very different values for rodent α: about two when both synonymous and intronic substitution rates are used (CHANG *et al.* 1994; CHANG and LI 1995). A comparison of autosomal and Y-linked substitution rates in rodents has led to an estimate of α close to one (MCVEAN and HURST 1997). What causes these differences, and what do such differences tell us about the applicability of MIYATA *et al.*'s (1987) method? We define three classes of explanations for the discrepancies between the X/A, Y/A, and X/Y estimates of α: statistical biases and artifacts, nonneutral evolution, and differences in mutation rate per germline replication.

Statistical biases and artifacts: The "errors" explanation supposes that there may be large standard errors in substitution rate estimates (SHIMMIN *et al.* 1993), and thus α values of two and ∞ may not be significantly different. An alternative "methodological" explanation suggests that the discrepancy between estimates of α may be due to different distance estimation methods yielding different substitution rates, as has been observed with other data sets (SMITH and HURST 1998b).

Nonneutral evolution: If synonymous substitutions in rodents are not selectively neutral, then different selective pressures on the different gene classes (X-linked, Y-linked, and autosomal) could lead to differences in K_S and thereby to discrepancies in estimates of α. For example, stronger selection could reduce synonymous substitution rates on the X-linked genes relative to the autosomal genes (SHIMMIN *et al.* 1993) because the X chromosome is exposed in males. Such exposure might strengthen selection against slightly deleterious synonymous mutations on the X chromosome relative to the

Corresponding author: Nick G. C. Smith, School of Biology and Biochemistry, University of Bath, Claverton Down, Bath BA2 7AY, United Kingdom. E-mail: n.smith@bath.ac.uk

autosomes, thereby reducing X-linked synonymous substitution rates. In contrast, synonymous substitution rates are less likely to be reduced on the Y chromosome despite the continual hemizygous expression of Y-linked genes. Y-linked genes are weakly expressed and mutations on most Y-linked genes might well be masked by their X-linked paralogs (Hurst and Ellegren 1998). Both reasons suggest that selection on the Y chromosome against slightly deleterious mutations need be no stronger than on the autosomes.

This problem of nonneutrality might be circumvented by using intronic, rather than synonymous, substitution rates. If intronic substitutions are neutral in mammals, as is generally thought, then estimates of α based on intronic substitution rates should be safer than estimates based on synonymous substitution rates (Shimmin *et al.* 1993). But alignment difficulties make the intronic substitution rate estimation difficult (Smith and Hurst 1998b), and selection seems to affect introns as well as silent sites in Drosophila (Bauer and Aquadro 1997).

Differences in mutation rate per germline replication: The X/A and Y/A estimates of α are based on nonhomologous comparisons while the Y/X estimates have been based on homologous comparisons, which are presumably more reliable. The "compositional" explanation proposes that base composition effects on substitution rates could explain the discrepancy between X/A and Y/X estimates of α (Shimmin *et al.* 1993), because base frequency differences can lead to different mutational biases (Morton and Clegg 1995; Morton *et al.* 1997).

DNA methylation status is known to affect mutation rates strongly, with methylated CpG's mutating to TpG's at 10–20 times the rate of unmethylated CpG's (Kendrew 1994). If X-linked genes are methylated less than autosomal or Y-linked genes, then a low mutation rate on the X would be expected. This "methylation" explanation differs from the compositional explanation in that the mutation rate depends on base modifications rather than the bases themselves.

The "mutation rate selection" explanation proposes that selection might favor different optimal mutation rates for the different gene classes. The mutation rate on the X chromosome may be selectively lowered relative to that on the autosomes due to the exposure of highly deleterious mutations in males (McVean and Hurst 1997). Selection to reduce the mutation rate on the Y chromosome is expected to be weak because there are few genes on the Y chromosome, of which many are weakly expressed and/or masked by X-linked paralogs (Hurst and Ellegren 1998). A reduction of substitution rates on the X chromosome would lead to discrepancies between X/A, Y/A, and X/Y estimates of α for both intronic and synonymous substitution rates. Note that a modifier of mutation rate could act via methylation levels or composition. Thus the three explanations in the "differences in mutation rate per germline replication" class (methylation, composition, and mutation rate selection) are not necessarily in competition.

The "nonneutral evolution" and differences in mutation rate per germline replication classes of explanation both predict that hemizygously expressed genes should have a lower substitution rate than diploid expressed genes. We test this hypothesis by asking whether imprinted genes have low K_S values. Genomically imprinted genes are those for which expression is dependent upon the sex of the parent from which they are derived (Efstratiadis 1994). When reasoning similar to that employed for X-linked genes is used, both the nonneutral evolution and mutation rate selection arguments are consistent with reduced K_S values for imprinted genes relative to autosomal genes. However, X-linked genes and imprinted genes are not expected to yield identical substitution rates because there are several differences between the two classes of genes: the proportion of time hemizygously expressed, the effective population size, and potential expression level, composition, methylation, and recombination rate differences.

MATERIALS AND METHODS

Selection of protein coding sequences: A list of 470 mouse/rat mRNA pairs, with HOVERGEN 19 used to confirm orthology (Duret *et al.* 1994), was obtained from Makalowski and Boguski (1998). Mouse chromosome locations and gene names were obtained from the Mouse Genome Database (MGD; http://mgd.hgmp.mrc.ac.uk/). A total of 297 pairs of autosomal mouse-rat orthologs were determined. MGD was used to obtain a list of X-linked mouse mRNA sequences excluding those in the pseudoautosomal region. Gapped BLAST (Altschul *et al.* 1997) at the NCBI (http://www.ncbi.nlm.nih.gov/) was used to determine rat X-linked orthologs. A total of 37 pairs of X-linked mouse-rat orthologs were determined.

Fifteen mouse/rat imprinted orthologs were determined, with all mouse genes given in the Mammalian Genetics Unit Database (http://www.mgu.har.mrc.ac.uk/), with the exceptions of *Gabrb3* and *Mas* (see below). Gapped BLAST was used to find rat orthologs and MGD to provide gene names. A total of 15 pairs of imprinted mouse-rat orthologs were determined. The imprinted status of *Gabrb3* has yet to be fully demonstrated; however, the evidence in favor is now fairly compelling (Saitoh *et al.* 1994; Culiat *et al.* 1995; Odano *et al.* 1996; Delorey *et al.* 1998; Meguro *et al.* 1997). The status of *Mas* is ambiguous, with two reports supporting imprinting (Villar and Pedersen 1994; Miller *et al.* 1997) and two failing to find evidence of imprinting (Riesewijk *et al.* 1996; Schweifer *et al.* 1997). Given that the likelihood of a false positive is probably less than that of a false negative, we consider that the balance of evidence is adequate to allow its incorporation.

Selection of intron sequences: A list of intronic and exonic mouse/rat pairs of confirmed orthology was obtained from Hughes and Yeager (1997). MGD was used to determine gene names and chromosomal location, which yielded 20 complete coding sequences and 70 introns. The X-linked mouse/rat pairs described above were searched for intron sequences, which gave three exon and five intron X-linked pairs.

Selection of sequences for Y/X comparison: Alignments used in the Y/X comparisons were obtained from a previous

study (McVEAN and HURST 1997), with both protein coding and intron alignments available for *Sry*, *Zfy*, and *Ube1y* (names as at MGD).

Sequence alignment: Alignments were performed using the GCG (GENETICS COMPUTER GROUP 1994) and EGCG (RICE 1997) sequence manipulation packages at HGMP (http://www.homepage.mrc.ac.uk/). FETCH was used to extract sequences from databases. GENETRANS was used to extract and combine exons automatically, while SEQED was used to extract introns manually. End-weighted PILEUP with the default gap penalties was used to perform exonic and intronic alignments. Exonic alignments were (if necessary) corrected by GAPFRAME to avoid frameshifts. In some cases, GAPFRAME could not correct the alignments. Then default PILEUP was used to produce protein alignments, and the program MRTRANS was used to recreate the DNA alignments from the protein alignments and the original DNA sequences (written by B. Pearson and available at HGMP). End-weighted CLUSTALW with the default gap penalties was also used to produce intronic alignments (THOMPSON *et al.* 1994), because it has been shown that different alignment packages with different default gap penalties produce significantly different intronic alignments (SMITH and HURST 1998b), and because there is no procedure for optimizing gap penalties (ALTSCHUL 1997).

Distance estimation: Several distance estimation methods were used. The default (and hopefully optimum) algorithmic estimation protocol used methods developed by MORIYAMA and POWELL (1997) with TAMURA's (1992) multiple hits correction method combined with LI's (1993) method for K_S and K_A estimation. Alternatively, KIMURA's (1980) correction for multiple substitutions was applied to reduce the estimation error. Four different types of substitution rate were estimated: the synonymous rate (K_S), the nonsynonymous rate (K_A), the synonymous rate at fourfold degenerate sites (K_4), and the intronic rate (K_I). Estimates of K_4 are less sensitive to methodology than estimates of K_S (LI 1993).

The methods of LI *et al.* (1985) with KIMURA's (1980) two parameter method for multiple substitution correction were also used to calculate K_S values. This enabled a more direct comparison with previous analyses that used such methods (WOLFE and SHARP 1993; MCVEAN and HURST 1997). Note that the method of LI *et al.* (1985) overestimates K_S by ~30% (LI 1993).

The maximum-likelihood package PAML (YANG 1997) was used to estimate substitution rates under a number of different assumptions. The program CODEML was used to estimate K_A and K_S under three settings: (1) "constant," a single rate for all sites; (2) "variable," a gamma distribution for variable substitution rates across sites; and (3) "correlated," a gamma distribution for variable substitution rates across sites and correlation of rates at adjacent codons. It should be noted, however, that the results obtained using the "variable" and "correlated" settings should be treated with caution since the gamma distribution may not apply to codon rates (Z. YANG, personal communication).

Calculating error limits of α estimates: The autosomal and X-linked synonymous and intronic substitution rates were compared to provide an estimate of α in a three-step procedure. First, substitution rate means and standard errors were calculated. Second, the X-to-autosome ratio of mean rates was calculated, and confidence limits for the ratio were calculated by adjusting both numerator and denominator by one standard error. Finally, MIYATA *et al.*'s (1987) equation of $K_X/K_A = (2/3)(2 + \alpha)/(1 + \alpha)$ was used to calculate both α and its confidence limits.

Codon usage bias: One measure of the codon usage bias of a gene is given by the effective number of codons (ENC; WRIGHT 1990). This statistic can vary from 20, with one codon used exclusively for each amino acid (high codon bias), to 61, with all synonymous codons used equally (low codon bias). Thus ENC is negatively correlated with codon usage bias.

We considered the possibility that ENC might be affected by both gene length and compositional bias. Random effects on smaller codon genes might reduce ENC, although it appears that ENC is highly robust to changes in gene length (COMERON and AGUADE 1998; MORIYAMA and POWELL 1998). Compositional bias makes codon usage less even, thereby reducing ENC (WRIGHT 1990). To account for these effects, a simulated ENC value was calculated for each gene. The codon position-specific base frequencies (*e.g.*, frequency of C at the first codon position) of the gene were used to create 1000 simulated gene sequences of the same length as the original gene. Mean ENC was calculated for each set of simulated sequence, then the simulated ENC was divided by the real ENC to give the statistic CUBRE (codon usage bias relative to expected). This way of treating the data allows the CUBRE values to positively correlate with codon usage bias. Both ENC and CUBRE were used because the two statistics provide alternative null hypotheses of codon usage. ENC assumes no mutational biases between bases and hence equal use of all codons. CUBRE assumes that compositional biases are solely the result of mutational biases rather than selection on base composition.

Codon usage heterogeneity: A test of codon usage bias that accounts for mutational biases that might be caused by adjacent bases has been developed by EYRE-WALKER (1991). Four nonoverlapping sets of codons are specified such that, within each set, all codons have the same base at the second codon position. For each set of codons, codon numbers are adjusted for fourth position base composition (*i.e.*, composition at the first base of the next codon). Then a chi-square test of heterogeneity is applied to determine whether the different amino acids within each set of codons show different codon usage patterns. Scaled chi-square values (similar to the codon usage statistic of SHIELDS *et al.* 1988) are obtained by dividing chi-square values by the number of codons within each set (after adjustment for fourth-position base composition).

Using Eyre-Walker's notation, the four sets of codons are C (alanine, threonine, proline, and fourfold codons of serine; all codons have C at the second position), AGA (glutamic acid, lysine, and glutamine; all codons have A at the second position and either G or A at the third position), ATC (tyrosine, histidine, aspartic acid and asparagine; all codons have A at the second position and either T or C at the third position), and GTC (cysteine and the twofold degenerate codons of serine; all codons have G at the second position and T or C at the third position). The unscaled and scaled chi-square values for the AGA test, for example, are denoted χ^2_{AGA} and $sc\chi^2_{AGA}$, respectively.

If genes are too short, then codon numbers are too small, and the values obtained from the tests will not be chi-square distributed. Thus only suitably long genes were analyzed, with the adjusted number of codons required to be greater than the number of different codons multiplied by five. Too few X-linked and imprinted genes were long enough, so only the results obtained from autosomal genes are considered. Of the mouse autosomal genes, the numbers of genes used in the different tests were as follows: C test, 118; AGA test, 177; ATC test, 91; and GTC test, 85. Of the rat autosomal genes, the numbers were as follows: C test, 128; AGA test, 184; ATC test, 97; and GTC test, 84.

For these tests to be capable of detecting selection for optimal codon usage, it is required that amino acids within the same set of codons differ with respect to third position base of the optimal codon or at least with respect to the degree to

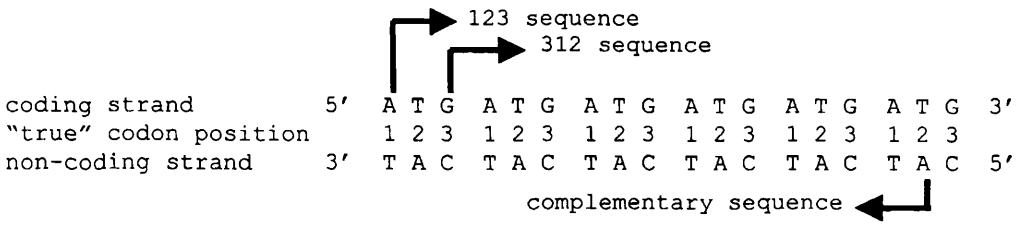


FIGURE 1.—Different sequences used in the analysis of codon usage heterogeneity (see MATERIALS AND METHODS). Both the 123 and 312 strands are read 5′ to 3′ on the coding strand. The 123 sequence starts at position 1 of the first codon, while the 312 sequence starts at the third position of the first codon. The complementary strand is read 5′ to 3′ on the noncoding strand from the second position of the first codon.

which the optimal codon is favored. The data from Drosophila, with most amino acids having the same optimal third-position base (see AKASHI 1994), suggest that Eyre-Walker's test may be weak.

EYRE-WALKER (1991) details how comparisons of the codon usage heterogeneities of different reading frames can provide information on the nature of selection acting to cause such heterogeneities. He defines the 123 sequence as the complete original sequence, the 312 sequence as the original sequence read 5′ to 3′ from the third base, and the complementary sequence as the complementary strand read 5′ to 3′ from the second base (see Figure 1). Selection for optimal codon usage should give bias on the 123 frame only, while selection on RNA structure should give bias on the 123 and 312 reading frames, and selection for DNA structure should give bias on both reading frames and the complementary sequence.

Compositional analysis: A number of compositional characters were measured: all four base frequencies at fourfold degenerate sites and overall CpG frequency. Predicted CpG frequencies were calculated using gene length and C and G overall base frequencies. In addition, CpG/TpG and CpG/CpG orthologous pairs where the C/T change is silent (*i.e.*, C/T at third codon position) were counted. From this analysis, silent CpG mutability, defined as the ratio of mutated (CpG/TpG) to unmutated (CpG/CpG), could be calculated. Methylated CpG dinucleotides are known to mutate to TpG at high rates (KENDREW 1994), and thus the silent CpG mutability provides a measure of germline methylation density.

Statistical methods: We performed nonparametric tests of statistical significance. The Mann-Whitney *U*-test was used to compare sets of data (such as the substitution rates of X-linked and autosomal genes). Rank correlation statistics followed by the z^* transformation suggested by Hotelling (SOKAL and ROHLF 1995), were used to examine the relationships between different gene characters. The Wilcoxon test was used to compare samples to expected values. As described above, we used chi-square tests to examine codon usage heterogeneity.

RESULTS

Statistical biases and artifacts

A number of different synonymous rate measures were used to provide estimates of α (Table 1). The finding that synonymous substitution rates are significantly lower for X-linked than autosomal genes holds for all measures (PAML correlated, $P = 0.032$; all other measures, $P < 0.0001$) and appears robust to the effects of outliers (data not shown). The eight different synonymous measures of α range from eight to ∞. In contrast with previous results (MIYATA *et al.* 1987; WOLFE and SHARP 1993; MCVEAN and HURST 1997), our dataset estimates α to be less than (but not significantly different from) ∞ for six out of the eight measures. Differences in methodology do not appear to lead to qualitative differences in α estimation. The upper confidence limits for α are very high (∞ for six out of eight measures), while the lower confidence limits range between three and seven.

Four intronic measures of α were obtained (Table 1). As for the synonymous data, X-linked rates are lower than those on the autosomes, although not significantly so. The multiple substitutions correction method appeared to make little difference, but the two different alignment protocols gave rather different estimates of α. PILEUP estimated α to be ~17 and gave a minimum α estimate of 1.7, which is close to the values obtained from X/Y comparisons. CLUSTALW yielded an α estimate of ∞ and a lower limit of 4.4, results that are more in keeping with the synonymous measures. These intronic results should be treated with caution because of the small sample size ($N = 5$ on the X chromosome).

Imprinted genes do have low K_S values

If the discrepancies in α estimates are solely due to statistical biases and artifacts and the other explanations are incorrect, then hemizygously expressed genes should have the same synonymous substitution rates as diploid-expressed genes. However, we find that a low K_S may be a general feature of haploid-expressed genes, both X-linked and imprinted: imprinted genes have both a significantly lower mean K_S ($P = 0.0046$) and a significantly lower mean K_4 ($P = 0.0179$) when compared with autosomal genes when the default algorithmic estimation method is used (see Figure 2). When maximum-likelihood methods are used, autosomal K_S is higher than imprinted K_S (constant $P = 0.0034$, variable $P = 0.0017$, and correlated $P = 0.033$; see Figure

TABLE 1

A number of different methodologies are used for derivations of α estimates with X/A comparisons

Rate	Method	X mean	A mean	X/A	Mean α	Min. α	Max. α
K_S	PBL T	0.125	0.178	0.7	18.5	5.9	∞
K_S	PBL K	0.124	0.175	0.708	15.2	5.4	∞
K_S	LWL K	0.158	0.223	0.708	15.1	5.4	∞
K_4	T	0.132	0.18	0.735	8.7	3.9	119
K_4	K	0.132	0.178	0.741	7.95	3.7	60
K_S	PAML constant	0.125	0.183	0.686	33.2	7	∞
K_S	PAML variable	0.149	0.321	0.464	∞	∞	∞
K_S	PAML correlated	0.437	0.832	0.526	∞	∞	∞
K_I	PILEUP T	0.156	0.221	0.703	17.5	1.7	∞
K_I	PILEUP K	0.155	0.221	0.704	16.7	1.7	∞
K_I	CLUSTALW T	0.137	0.233	0.587	∞	4.4	∞
K_I	CLUSTRALW K	0.137	0.232	0.589	∞	4.3	∞

The rows show the steps in calculating α and its confidence limits (see also RESULTS section). The X chromosome and autosomal means are used to calculate the ratio of means (X/A). Then the ratio (and its upper and lower limits) is used to estimate α (and its lower and upper limits). The columns give different methodologies. K_S, K_4, and K_I are substitution rates at synonymous, fourfold degenerate, and intronic sites, respectively. The different methods are abbreviated as follows: PBL, LI (1993); LWL, LI *et al.* (1985); K, KIMURA (1980); T, TAMURA (1992). See MATERIALS AND METHODS for details of the different PAML methods. The alignment program (PILEUP or CLUSTALW) is given fir the K_I data to demonstrate the strong effect of alignment protocol.

3). Similarly, autosomal K_S is higher than X-linked K_S when such methods are used (constant $P < 0.0001$, variable $P < 0.0001$, and correlated $P = 0.031$; see Figure 3). Both the nonneutral evolution and mutation rate selection arguments predict both X-linked and imprinted genes to have lower K_S rates than autosomal genes (see Introduction).

Nonneutral evolution of silent sites is not supported

We examine two issues concerning the proposition that nonneutral evolution of third-site mutations can account for the low K_S values seen in X-linked and imprinted genes. First we provide codon usage analyses to ask whether there is any evidence that selection has affected codon usage in rodents. Second, we ask whether, in principle, nonneutral evolution is adequate as a possible explanation.

Codon usage data: Although there exists a wealth of evidence to suggest that codon usage is selectively driven in Drosophila and bacteria (see LI 1997), several lines of evidence suggest that in mammals silent sites are neutral (see MCVEAN and HURST 1997). The "silent site selection" explanation belongs to the nonneutral evolution class of theories (see Introduction) and requires silent sites to be selectively constrained. We performed two sets of tests on rodent sequences to ask whether there is any evidence that codon usage is af-

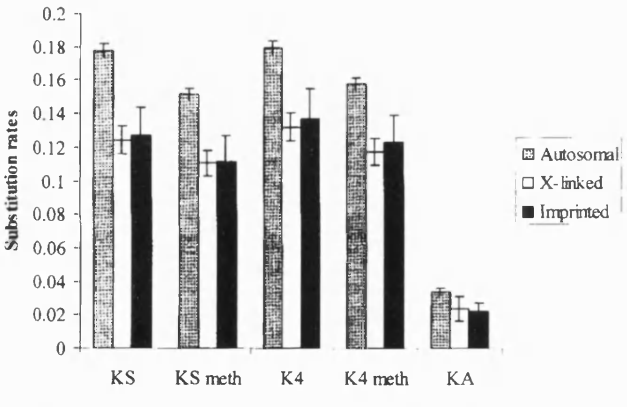


FIGURE 2.—Autosomal, X-linked, and imprinted genes compared with respect to synonymous substitution rate (K_S), nonsynonymous substitution rate (K_A), and fourfold substitution rate (K_4). The error bars give standard errors. The methods of TAMURA (1992) and LI (1993) were used to calculate substitution rates. "meth" refers to rates estimated by excluding from consideration CpG ⇔ TpG substitutions at silent sites.

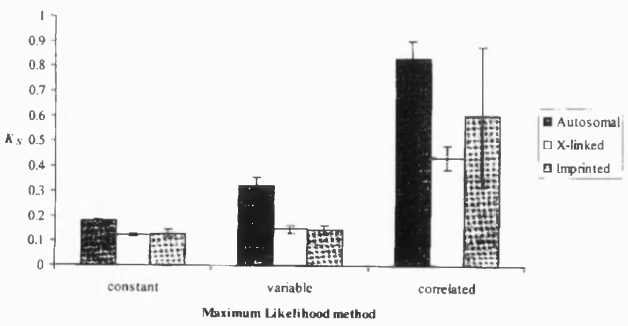


FIGURE 3.—Autosomal, X-linked, and imprinted genes compared with respect to synonymous substitution rates (K_S) obtained when different settings of the maximum-likelihood program PAML were used (see MATERIALS AND METHODS).

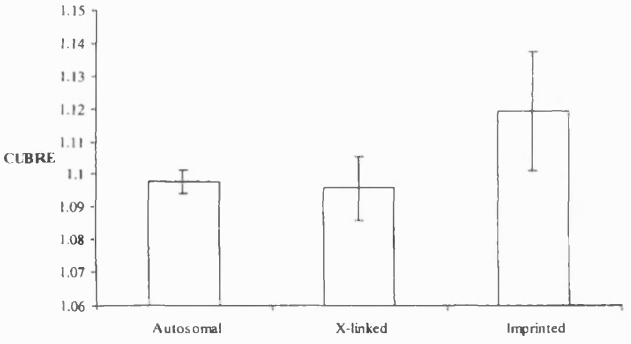


FIGURE 4.—Autosomal, X-linked, and imprinted genes compared with respect to CUBRE (codon usage bias relative to expected; see MATERIALS AND METHODS).

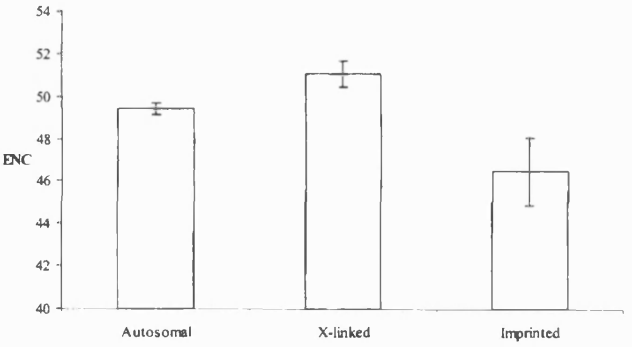


FIGURE 5.—Autosomal, X-linked, and imprinted genes compared with respect to ENC (effective number of codons; see MATERIALS AND METHODS).

fected by selection and whether codon usage bias can explain the variation that we see in K_S.

CUBRE and ENC analysis: If synonymous codon usage were selectively driven, then one would expect to see a correlation between K_S and CUBRE and between K_S and ENC. If the synonymous codon usage of a gene is under strong selection, then CUBRE will be high and ENC low (high codon bias) because only optimal codons will be used, and K_S will be low because synonymous changes (from an optimal codon to a nonoptimal codon) will be highly constrained. Conversely, if the synonymous codon usage of a gene is under weak selection, then CUBRE will be low and ENC high (low codon bias) and K_S high. Such correlations have provided evidence for selection on codon bias in Drosophila (SHARP and LI 1989).

The predicted relationships between CUBRE and ENC values and K_S were tested in two different ways. First, if mean K_S is significantly higher for autosomal genes than for X-linked genes, do mean CUBRE and mean ENC show the expected relationships (significantly higher and lower, respectively, on the X chromosome)? Second, is there a significant correlation between K_S and CUBRE and between K_S and ENC among either autosomal or X-linked genes?

Both K_S and K_4 are significantly higher for autosomal than X-linked genes under the default algorithmic estimation method ($P < 0.0001$ in both cases, see Figure 2). K_S is also significantly higher for autosomal than X-linked genes under maximum-likelihood estimation (constant $P < 0.0001$, variable $P < 0.0001$, and correlated $P = 0.03$).

In conflict with the predictions of the silent site selection argument, mean CUBRE is lower ($P = 0.99$; see Figure 4) and mean ENC is higher ($P = 0.09$; see Figure 5) on the X-linked genes. No significant correlations were found between K_S and CUBRE (autosomal and X-linked, $P > 0.1$ and $P > 0.1$) or K_S and ENC ($P > 0.1$ and $P > 0.05$).

Considering imprinted genes allows a further test of the silent site selection argument. Given that imprinted genes have significantly lower synonymous substitution rates than autosomal genes (see above), the silent site selection argument predicts that imprinted genes should show greater codon usage bias; *i.e.*, imprinted genes should have a higher mean CUBRE and lower mean ENC than autosomes. These predictions hold for both CUBRE ($P = 0.35$) and ENC ($P = 0.09$), although the results are not significant.

Codon usage heterogeneity: Although the results of the codon bias analysis presented above provide no evidence that silent sites in rodents are under selection, CUBRE is significantly greater than unity in each of the three classes of autosomal, X-linked, and imprinted genes ($P < 0.001$ in all three cases). This result means that either silent sites are subject to selective pressures (and for some reason our previous tests did not have the power to demonstrate selection) or the scheme used to produce the expected ENC values is incorrect. The latter explanation seems plausible because simulated ENC values were calculated using only position-specific base frequencies and gene length. It is known that neighboring bases can affect mutation rates (BULMER 1986), and thus simply combining position-specific base frequencies is unlikely to give reliable codon frequency estimates.

To control for mutational biases caused by adjacent bases, the codon usage heterogeneity tests developed by EYRE-WALKER (1991) were applied to those autosomal genes that fulfilled the gene length criteria (see MATERIALS AND METHODS). Restricting our analysis to autosomal genes, we aimed to determine whether the result of mean CUBRE greater than one is a methodological artifact.

For each of the four different tests (χ^2_{GTC}, χ^2_C, χ^2_{AGA}, and χ^2_{ATC}), two statistics were calculated: the proportion of genes showing significant heterogeneity at the 5% level and the overall χ^2 for all the genes. For both mouse and rat, only one of the overall χ^2 tests, overall χ^2_{AGA}, showed significant heterogeneity ($P < 0.001$ in both species). Only the rat χ^2_{AGA} test gave significantly more genes with significant bias than expected by chance ($P <$

0.025 with Yates' continuity correction), though the mouse χ^2_{AGA} test showed a similar tendency ($P < 0.1$).

Those genes that showed significant heterogeneity in any of the four tests in either mouse or rat were classed as the high χ^2 group, while the remaining genes with nonsignificant heterogeneities in both mouse and rat for all four tests formed the low χ^2 group. If the silent site selection argument is correct, then the high χ^2 group (more codon bias) should have lower K_S (silent sites more constrained) than the low χ^2 group. In conflict with this prediction, the high χ^2 group had a higher mean K_S than the lower χ^2 group, though the two means were not significantly different ($P = 0.21$).

The χ^2_{AGA} test was repeated with analysis of the 123, 312, and complementary sequences (see MATERIALS AND METHODS). All autosomal genes that fulfilled the χ^2_{AGA} test length criteria for the 123, 312, and complementary sequences were included, thereby increasing the sample size. For both mouse and rat, overall χ^2_{AGA} was significant for the 123 sequence ($P < 0.001$ for both species) and complementary sequence ($P < 0.001$ for both species) but not for the 312 sequence ($P > 0.25$ for both species).

As explained above, the silent site selection argument predicts a negative correlation between codon bias and K_S. The four tests of heterogeneity were used to test this prediction, with scaled values used to control for gene length (see MATERIALS AND METHODS). None of the four scaled heterogeneities correlated significantly with K_S ($P > 0.1$ in all four cases).

Nonneutral evolution, the problem of advantageous recessives, and the K_A-K_S correlation: One assumption of the notion that a low K_S in imprinted and X-linked genes might be the result of nonneutral evolution is that advantageous recessive silent mutations are rare. If advantageous recessives were common, then the exposure of X-linked and autosomal genes could lead to them having higher substitution rates than autosomal genes (CHARLESWORTH *et al.* 1987), a result that would conflict with the low K_S values reported here.

We cannot test the frequency of advantageous recessive mutations directly. However, we can ask whether nonsynonymous substitutions show any evidence for the presence of advantageous recessive mutations. If purifying selection had acted on synonymous mutations to reduce X-linked K_S, one would expect an even greater reduction in X-linked K_A if nonsynonymous mutations were also under purifying selection, because nonsynonymous mutations are likely to be affected by stronger selection than synonymous ones.

K_A is significantly higher on the autosomes than on the X chromosome ($P = 0.003$; see Figure 2). Although the relationship is not significant, K_A is higher on autosomal genes than imprinted genes ($P = 0.52$; see Figure 2). These K_A relationships cannot, however, be considered independently of the K_S relationships because of the well-known K_A-K_S correlation (*e.g.*, MOUCHIROUD *et al.* 1995). Therefore, following MCVEAN and HURST (1997), we ask if X-linked K_A values are lower than one might expect given the low K_S values on the X chromosome. If the dominant mode of selection on the X were stabilizing selection then X-linked genes should have low K_A values given their K_S values. If instead directional selection on advantageous recessives were common then both might have K_A values higher than or comparable with autosomal genes with comparable K_S values.

We compared two rank orders, one for K_S and one for K_A, both of which gave the ranks of X-linked genes among autosomal genes. The K_A ranks were higher than the K_S ranks but not significantly so ($P = 0.14$). Thus we can reject the notion that X-linked K_A values are lower than expected on the basis of K_S values, which is a result that provides evidence against nonneutral evolution reducing the K_S on the X chromosome.

A similar analysis was applied to the imprinted genes. As for the X-linked genes, if the dominant mode of selection affecting imprinted genes is stabilizing selection then imprinted genes should have low K_A values given their K_S values. For imprinted genes among autosomal genes there is a tendency toward higher K_A ranks ($P = 0.08$). Consistent with the similarly high K_A ranks of both X-linked and imprinted genes, the comparison of imprinted and X-linked genes suggests no difference between the two classes ($P = 0.69$ for X-linked among imprinted and $P = 0.56$ for imprinted among X-linked). This result is consistent with elevated K_A ranks being an effect of hemizygous expression. If the hemizygously expressed genes are pooled, then the K_A ranks among the autosomal genes are significantly higher than the K_S ranks ($P = 0.041$).

The X and autosomes differ in sequence composition

The composition and methylation explanations for the discrepancies between X/Y, Y/A, and X/A estimates of α both invoke differences in the mutation rate per germline replication, which are the result of sequence differences. We tested the idea that the observed differences between X-linked, imprinted, and autosomal synonymous substitution rates might be due to sequence differences.

Most of the compositional features considered differ significantly between the autosomal genes and the X-linked genes (see Table 2). Most notably the mean GC4% on the X chromosome was only 53% compared with 61% on autosomes. This effect cannot account for all the variation in K_S, because imprinted genes, which also have a low K_S, have a GC4% of 65%, which is higher than that of the autosomal genes. No compositional feature differs significantly between the autosomal and imprinted genes (Table 2). We do find, however, that imprinted genes tend to be more GC rich than autosomal genes ($P = 0.053$). The fact that the difference is close to significance may help to explain the different

TABLE 2

Compositional statistics for autosomal (A), X-linked (X), and imprinted (I) genes

	Autosomal	X-linked	Imprinted	P (A = X)	P (A = I)
A4	0.184	0.204	0.164	0.06	0.2
C4	0.345	0.307	0.359	0.006	0.43
G4	0.267	0.221	0.289	<0.0001	0.22
T4	0.204	0.268	0.187	<0.0001	0.28
GC4	0.612	0.529	0.649	0.0001	0.053
TpG/CpG	1.184	1.502	0.688	0.52	0.075
CpG O/E	0.462	0.404	0.485	0.0037	0.92

Means are given, along with the Mann-Whitney P values for the AX and AI comparisons. A4, C4, G4, T4, and GC4 are the base frequencies at fourfold degenerate sites. TpG/CpG gives the ratio of silent Cpg ⇔ TpG substitutions to conserved silent CpG sites (see MATERIALS AND METHODS). CpG O/E is the observed-over-expected ratio of CpG sites (equation for calculating expected values given in MATERIALS AND METHODS).

results obtained by previous compositional comparisons of imprinted genes and autosomal genes (NEUMANN *et al.* 1995; MCVEAN *et al.* 1996).

In addition to compositional differences, we have found evidence of methylation differences (see MATERIALS AND METHODS). On the X chromosome not only are C's and G's rare, but also the frequency of CpG's (with C at a silent site), when controlled for GC%, is lower than that found on the autosomes ($P = 0.037$; see Table 3). But is there evidence that such differences lead to synonymous substitution rate differences?

Autosomal genes were used to examine correlations between sequence characters and synonymous substitution rates. No significant correlations were observed between K_4 and any of the fourfold base frequencies ($P > 0.5$ in all four cases). The number of silent CpG ⇔ TpG substitutions correlates positively with K_S ($P < 0.001$), which is not surprising given that such changes contribute directly to the number of synonymous substitutions. However, the ratio of silent CpG ⇔ TpG changes to the number of conserved silent CpG sites, which gives an indication of the mutability of silent CpG sites (see MATERIALS AND METHODS), also correlates strongly with K_S ($P < 0.001$). In other words, the synonymous substitution rate at a class of site known to be strongly influenced by methylation status correlates strongly with the whole gene synonymous substitution rate. Such an effect has already been shown to explain at least some of the K_S variation within *Igf2r*, one of the imprinted genes in this study (SMITH and HURST 1998a). The ratio of observed to expected CpG (see MATERIALS AND METHODS) also shows a positive correla-

TABLE 3

Estimates of α obtained using a variety of methodologies that use all three comparisons of X/A, Y/A, and X/Y

	K_S	K_S meth	K_4	K_I	PAML K_S constant	PAML K_S variable	PAML K_S correlated
X mean	0.125	0.111	0.132	0.137	0.125	0.149	0.4377
A mean	0.178	0.152	0.18	0.233	0.182	0.321	0.8327
Y mean	0.183	0.171	0.238	0.194	0.182	0.238	0.1997
Zfx	0.0833	0.0785	0.0867	0.144	0.0771	0.0771	0.335
Xfy	0.208	0.208	0.232	0.231	0.205	0.246	0.354
Ube1x	0.0861	0.0784	0.0796	0.167	0.085	0.0876	0.0335
Ube1y	0.138	0.119	0.223	0.175	0.121	0.13	0.125
α X/A	18.5	9.59	8.69	∞	33.2	∞	∞
α X/Y	1.93	2.12	2.99	1.79	1.88	2.28	0.358
α Y/A	1.07	1.29	1.95	0.711	0.994	0.588	0.136
α *Zfx/Zfy*	9.96	15.2	16.7	2.31	15.6	∞	1.09
α *Ube1x/Ube1y*	2.31	2.06	28.8	1.07	1.82	1.95	∞

Substitution rates (K_S, synonymous; K_4, fourfold degenerate; and K_I, intronic) were estimated either by maximum likelihood using the program PAML (see MATERIALS AND METHODS for the settings used) or by the algorithmic methods of TAMURA (1992) and LI (1993). The "K_S meth" data were obtained by excluding CpG ⇔ TpG substitutions at silent sites from consideration. The default settings of the CLUSTALW were used to prepare intronic alignments. Values of α for the X/A, X/Y, and Y/A comparisons were calculated according to MIYATA *et al.*'s (1987) equations.

tion with K_S ($P < 0.02$). This positive correlation is surprising, given that heavily methylated genes might be expected to have both high rates of silent substitution and low ratios of observed over expected CpG (both due to the methylation-induced mutation of CpG dinucleotides). It might be that methylated genes are under particularly strong stabilizing selection for some reason.

Could these links between methylation and K_S provide an explanation for the variation in K_S between genes? Silent site methylation-induced mutability is higher on the X chromosome than on the autosomes, which is the wrong direction for explaining the low K_S of X-linked genes. To test the methylation explanation more rigorously, K_S and K_4 values were calculated for all gene pairs with all silent CpG ⇔ TpG changes ignored. Autosomal K_S remained significantly higher than both X-linked K_S ($P < 0.0001$) and imprinted K_S ($P < 0.01$). Autosomal K_4 was significantly higher than X-linked K_4 ($P < 0.0001$) but was not significantly higher than imprinted K_4 ($P = 0.058$). Furthermore, the discrepancies between X/A and X/Y estimates of α remain even after removal of such methylation-induced changes (see below and Table 3).

X/Y and Y/A estimates of α

To reject an explanation for the discrepancies between estimates of α it is not enough to simply show that the explanation is unable to yield X/A α estimates close to the values of about two given by X/Y comparisons. It must also be demonstrated that the explanation is unable to raise the α estimates provided by X/Y and Y/A comparisons. Therefore we have used Y-linked sequences to give X/Y and Y/A estimates of α. The methodological explanation can be tested by observing the impact of changes in methodology on the relative α estimates of the X/A, X/Y, and Y/A comparisons.

Using mean chromosomal class substitution rates, the relative values of α estimates are conserved across methodologies (see Table 3), which suggests a rejection of the methodology explanation. For both algorithmic estimates of K_S, K_4, and K_I rates and for maximum-likelihood estimates of K_S, the X/A comparisons give much higher estimates of α than the X/Y and Y/A comparisons. However, the homologous X/Y comparisons of *Zfx/Zfy* and *Ube1x/Ube1y* do yield some α estimates greater than 2 (Table 3). The K_S *Zfx/Zfy* comparison gives an α estimate of 10. A previous study found $\alpha = \infty$ in this comparison (Shimmin *et al.* 1994), and although we concur with the result of a high α estimate, we disagree with the conclusion of the authors of that study that such a result was due to silent site constraints affecting *Zfx* being relaxed on *Zfy*. *Zfx* shows no evidence of codon usage bias [CUBRE = 1.017; *c.f.*, mean X-linked CUBRE = 1.0958; furthermore, none of Eyre-Walker's (1991) codon usage tests showed significant heterogeneity], and therefore silent sites on *Zfx* are unlikely to be under selective constraints. Using K_4 rates, both the *Zfx/Zfy* and *Ube1x/Ube1y* comparisons give high α estimates. This appears to be due to K_4 being greater than K_S on the Y chromosome, presumably due to a low rate at twofold degenerate sites.

The methylation explanation predicts that differences between chromosomes in methylation density could lead to differences between chromosomes in rates. Although this explanation does not appear to be able to reduce the X/A estimate of α to about two, might it be able to increase the X/Y and Y/A α estimates? To test this hypothesis we recalculated K_S values while ignoring CpG ⇔ TpG changes at silent sites (see Table 3). The X/A estimate decreased and the X/Y and Y/A estimates increased, but there remained large discrepancies between α estimates. The *Ube1x/Ube1y* comparison gave a lower α estimate on the removal of methylation-induced changes, but the *Zfx/Zfy* α estimate increased to a value actually above the X/A estimate.

Even if selection does act to reduce mutation rates on the X chromosome, it does not then automatically follow that mutation rates should also be selectively reduced on the Y chromosome (see Introduction). In keeping with this analysis K_S, K_4, and K_I are not nearly as low on the Y chromosome as they are on the X chromosome, although the sample size is very low (see MATERIALS AND METHODS).

DISCUSSION

We considered three explanations for the discrepancies between X/A, Y/A, and Y/X estimates of α (see Introduction).

Statistical biases and artifacts: Are the errors in estimates of α so large that values of 2 and ∞ are not significantly different? Or was it the use of biased distance estimation measures in previous analyses that led to the discrepancy in α estimates? Using both synonymous and intronic substitutions and a variety of methodologies, we used the X/A comparison to obtain estimates of α (see Table 1). The intronic and synonymous estimates differ with respect to expected value of α (finite but large for synonymous, infinite for intronic), but the large errors mean that both estimates provide 95% confidence intervals for α between 0.5–7 and ∞ when a range of different methods is used. These lower limits approach (and for the PILEUP intronic estimate fall below) estimates of α of ~2 obtained from Y/X comparisons (intronic and synonymous rates show similar results). However, we do not feel it is appropriate to calculate a statistic for the difference between the X/A and Y/X estimates of α, because the confidence limits previously obtained for Y/X estimates (Chang *et al.* 1994; Chang and Li 1995) are of a different nature than the confidence limits we have obtained.

The confidence limits provided in this study are based on the variation observed in synonymous substitution

rates between many genes (37 X-linked and 297 autosomal genes), which thereby reduce any bias caused by substitution rate variation (SHIMMIN *et al.* 1993). In contrast, studies that use Y/X comparisons to estimate α have ignored variation in substitution rates between genes, and have instead used the errors inherent in estimating substitution rates (CHANG *et al.* 1994; CHANG and LI 1995). This method might be defended on the basis that homologous genes should show little variation in substitution rates, but an assumption of absolutely no variation in rates between homologous genes (other than that caused by rate estimation) seems unlikely to be correct. Thus we regard the Y/X 95% confidence intervals of 1.0–3.2 (CHANG *et al.* 1994) and 1.0–3.9 (CHANG and LI 1995) as perhaps too narrow.

Because high estimates of α are obtained when both biased and unbiased methods of K_S and K_I estimation are used, the methodological explanation can be rejected. But unless X/A, Y/A, and X/Y estimates of α can be statistically compared, it is impossible to discount fully the errors argument. That only one of nine X/A α estimates has a 95% confidence limit under two (and that estimate was based on the smallest sample) does suggest rejection of the errors hypothesis. Furthermore, estimates of α that use the X/A, X/Y, and Y/A comparisons with the three distance measures of K_S, K_4, and K_I all yielded the same result of $\alpha_{X/A}$ greater than $\alpha_{X/Y}$ and $\alpha_{Y/A}$, which supports rejection of both the errors and methodological arguments.

It should be noted that the relatively small expansion in dataset size from MCVEAN and HURST's (1997) study to the present one appears to have caused a considerable increase in the K_S estimate of X/A. McVean and Hurst obtained an X/A ratio of 0.62. Even when the same methods as McVean and Hurst's are used, the dataset used here still gives an X/A ratio of 0.708. If further expansions in dataset size lead to similar increases in the X/A ratio, the X/A and Y/X estimates of α may converge further. Further evidence for larger datasets yielding higher X/A ratios comes from the human and rodent comparison. With 35 autosomal and only 4 X-linked genes, MIYATA *et al.* (1987) obtained an X/A ratio of 0.60, while a recent study of 559 autosomal and 71 X-linked genes (TERADA *et al.* 1997) gave an X/A ratio of 0.78.

Nonneutral evolution: The discrepancies between synonymous X/A, Y/X, and Y/A estimates of α may be due to stronger selection against silent mutations on the X chromosome (for further details see Introduction). If this is a necessary and sufficient explanation of α estimate discrepancies, then two requirements must be fulfilled. The first requirement is that there be no α estimate discrepancy when intronic data are used. The second requirement is that silent substitutions must be selectively constrained.

The initial requirement seems to be contravened by the intronic X/A, X/Y, and Y/A estimates of α. X/A gave $\alpha = \infty$, while X/Y and Y/A both gave α less than three (Table 3). However, it is not possible to determine whether the discrepancies in α estimates are significant.

Concerning the second requirement, our codon usage analyses provide no support for the hypothesis that silent substitutions in rodents are selectively constrained. Neither X-linked nor imprinted CUBRE is significantly higher than autosomal CUBRE, despite X-linked and imprinted K_S being significantly lower than autosomal K_S. Furthermore, neither autosomal nor X-linked genes demonstrate significant negative correlations between K_S and CUBRE.

Analysis of codon usage heterogeneities for autosomal genes reveals that only one of the four codon sets (A^G_A) shows significant levels of heterogeneity. Comparisons of A^G_A tests for different reading frames provide evidence that the significant heterogeneity for the 123 sequences is not the result of selection, because significant heterogeneity is found for the 123 and complementary sequences but not for the 312 sequences. Such results are similar to those of EYRE-WALKER (1991, p. 447), and we concur with this author's appraisal that "*It is difficult to think of any form of selection that would lead to significant bias appearing on the complementary strand and in other reading frames, but which is attenuated in the 312 reading frame.*"

Selection for optimal codon usage should give bias on the 123 frame only, while selection on RNA structure should give bias on the 123 and 312 reading frames, and selection for DNA structure should give bias on both reading frames and the complementary sequence. One possible explanation for the attenuation of bias in the 312 reading frame is the interaction of neighboring base mutational biases and longer range mutational biases (EYRE-WALKER 1991). The latter class of biases can act at distances of several bases (BULMER 1990), and are thus not accounted for by the method of second- and fourth-codon-position correction employed here. The base position examined for codon usage bias for both the 123 reading frame and the complementary sequence is the "true" third position (see Figure 1). Because the neighboring nucleotides (true first and second positions) are constrained, mutational biases will be able, over time, to generate significant heterogeneities at third positions. But the base position examined for codon usage bias for the 312 sequence is the true second position, which has the true first and third positions for neighbors. Because the third-codon position is only weakly constrained and thus rapidly evolving, mutational biases may be unable to build up sufficiently before the third site, and thus the direction of bias, changes.

The lack of significant heterogeneity in three tests out of four, the demonstration that the A^G_A bias is unlikely to be the result of selection for codon usage, and the fact that so few of the genes showed significant bias when considered individually all argue against the nonneutral

evolution theory. Furthermore, no negative correlation between codon bias and K_S was observed, and the "high bias" group of genes did not have a significantly lower K_S than the "low bias" group of genes.

Furthermore, it is unclear whether nonneutral evolution at third sites can explain the low K_S values that we observe. If the silent site selection explanation is correct, and if nonsynonymous mutations are as likely to be deleterious as synonymous ones, then the K_A values for the X-linked and imprinted genes should be lower relative to the autosomal genes than their K_S values. In fact, for both X-linked and imprinted genes, K_A values are higher than would be expected on the basis of their K_S values but not to a significant extent.

Differences in mutation rate per germline replication: *Composition and methylation explanations:* If the composition or methylation arguments are to explain low K_S on the X chromosome, there must exist significant sequence differences between the X chromosome and the autosomes, and those compositional characters that differ significantly must be shown to correlate (and in the required direction) with K_S. The former requirement is fulfilled by most of the compositional characters investigated (overall and fourfold base frequencies and CpG statistics); but only one sequence character was observed to correlate significantly with K_S or K_4. This was silent CpG mutability, which can be interpreted as a measure of methylation density.

Methylation at CpG sites appears to have a large effect on mutation rates. CpG ⇔ TpG substitutions constitute a reasonably large proportion of silent substitutions (the autosomal, X-linked, and imprinted estimates of K_S fell by 15, 11, and 12%, respectively, when such substitutions were ignored), and methylation levels can alter mutability without affecting the protein-coding sequence of a gene. Unfortunately for the methylation theory, X-linked genes actually showed greater silent CpG mutability than autosomal genes, and the removal of CpG ⇔ TpG substitutions from estimates of K_S had little effect on the significant difference between X-linked and autosomal K_S means. Such adjusted K_S values provide a lower X/A estimate of mean α than the unadjusted K_S values, although X/Y and Y/A estimates of α remain much lower than the X/A estimate.

It is surprising that CpG mutability is higher for X-linked genes than for autosomal genes, when one considers that male germline DNA is far more heavily methylated than female germline DNA (Monk *et al.* 1987). For example, Ketterling *et al.* (1993) found a male-to-female mutation bias of 11 when considering CpG dinucleotides in the factor IX gene. The reduction in the X/A estimate of α upon removal of silent CpG ⇔ TpG substitutions does fit in with a higher male bias for such mutations than for other sorts of point mutations.

Methylation status appears to be a crucial mechanism in genomic imprinting, and the fact that many imprinted genes possess CpG-rich regions implies that imprinted genes are relatively unmethylated in the germline (Constancia *et al.* 1998). Such a methylation profile is consistent with low K_S values for imprinted genes. Indeed, although the differences are not significant, the imprinted genes show a lower silent CpG mutability and a higher observed-over-expected CpG ratio (indicative of lower overall CpG mutability) when compared to the autosomal genes. However, after the removal of CpG ⇔ TpG substitutions imprinted genes still have a significantly lower K_S than autosomal genes (Figure 2), and so such substitutions cannot provide a complete explanation of the low K_S of imprinted genes.

It is impossible to refute conclusively either the composition or methylation explanations because in principle almost any sequence character could affect mutation rates, and one cannot examine every possibility.

Mutation rate selection explanation: The mutation rate selection explanation is supported by some circumstantial evidence. Both intronic and synonymous estimates of α appear to show discrepancies between the X/A and Y/X comparisons. Imprinted genes show the low K_S predicted on the basis of their hemizygous expression. The finding that K_S on the Y is higher than on the X is hardly supportive, but at least the result is consistent with selection on mutation rates.

Detailed analysis of imprinted genes provides further suggestive evidence in favor of the existence of modifiers of mutation rates. If imprinted genes have low K_S because of modifiers, then clusters of imprinted genes should have lower K_S than imprinted genes on their own, because the selective pressure favoring modifiers of mutation rate should increase with the number of genes affected by the modifier. So the more imprinted genes there are in a cluster, the better the chance of a modifier of the mutation rate becoming fixed. All of the imprinted genes in the dataset exist in clusters except for *Htr2a* (Yang *et al.* 1992), *Ins1* (Wentworth *et al.* 1986), and *Rasgrf1* (Plass *et al.* 1996). Nonclustered imprinted genes have higher K_S than clustered imprinted genes, and although the difference is not significant ($P = 0.22$) this may well be a result of limited sample size. We feel this result is suggestive and worthy of further analysis.

The status of Miyata's method for assessing α: How do the above results relate to the validity of Miyata *et al.*'s (1987) proposal that the male-to-female mutation rate ratio (α) can be estimated by comparing the substitution rates of X-linked, Y-linked, and autosomal genes? Their method (hereafter Miyata's method) depends on two assumptions. The first assumption is that mutation rates are, at least in principle, measurable. For molecular evolutionary methods this requires the presence of neutral sites and the use of appropriate multiple substitution correction methods. The second assumption is that all of the differences between the mutation rates of genes are because of the amount of time spent in the two germlines.

The statistical biases and artifacts explanation does not affect the logic underpinning Miyata's method but is certainly unsatisfying and probably incorrect. The nonneutral evolution hypothesis impinges on the first assumption that mutation rates are measurable. The evidence we have presented against this hypothesis lends credence to the first assumption of Miyata's method.

The differences in mutation rate per germline replication explanations suggest violations of the second assumption of Miyata's method. Different mutation rates on different chromosomes might be the result of different patterns of base composition or methylation or hemizygosity rather than the result of the amount of time spent in the two germlines. Neither the composition nor methylation explanations can be rejected (but given that both theories appear unfalsifiable, nothing can be read into this), and the evidence in favor of the mutation rate selection argument is only circumstantial. However, we consider the finding of a connection between low K_S and hemizygosity to be best interpreted as a violation of the second assumption of Miyata's method. Therefore we feel that caution should be exercised when interpreting estimates of α obtained when Miyata's method is used.

The authors thank Nick Britton, John Barrett, Gil McVean, Adam Eyre-Walker, Etsuko Moriyama, and two anonymous referees.

LITERATURE CITED

Akashi, H., 1994 Synonymous codon usage in Drosophila melanogaster—natural selection and translational accuracy. Genetics **136:** 927–935.

Altschul, S. F., 1997 Sequence comparison and alignment, pp. 137–168 in *DNA and Protein Sequence Analysis*, edited by M. J. Bishop and C. J. Rawlings. Oxford University Press, Oxford.

Altschul, S. F., T. L. Madden, A. A. Schaffer, J. H. Zhang, Z. Zhang *et al.*, 1997 Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. **25:** 3389–3402.

Bauer, V. L., and C. F. Aquadro, 1997 Rates of DNA sequence evolution are not sex-biased in Drosophila melanogaster and D-simulans. Mol. Biol. Evol. **14:** 1252–1257.

Bulmer, M., 1986 Neighboring base effects on substitution rates in pseudogenes. Mol. Biol. Evol. **3:** 322–329.

Bulmer, M., 1990 The effects of context on synonymous codon usage in genes with low codon usage bias. Nucleic Acids Res. **18:** 2869–2873.

Chang, B. H. J., and W.-H. Li, 1995 Estimating the intensity of male-driven evolution in rodents by using X-linked and Y-linked *Ube-1* genes and pseudogenes. J. Mol. Evol. **40:** 70–77.

Chang, B. H. J., L. C. Shimmin, S. K. Shyue, D. Hewett-Emmett and W.-H. Li, 1994 Weak male-driven molecular evolution in rodents. Proc. Natl. Acad. Sci. USA **91:** 827–831.

Charlesworth, B., J. A. Coyne and N. H. Barton, 1987 The relative rates of evolution of sex chromosomes and autosomes. Am. Nat. **130:** 113–146.

Comeron, J. M., and M. Aguade, 1998 An evaluation of measures of synonymous codon usage bias. J. Mol. Evol. **47:** 268–274.

Constancia, M., B. Pickard, G. Kelsey and W. Reik, 1998 Imprinting mechanisms. Genet. Res. **8:** 881–900.

Culiat, C. T., L. J. Stubbs, R. P. Woychik, L. B. Russell, D. K. Johnson *et al.*, 1995 Deficiency of the beta-3 subunit of the type-a gamma-aminobutyric-acid receptor causes cleft-palate in mice. Nat. Genet. **11:** 344–346.

Delorey, T. M., A. Handforth, S. G. Anagostaras, G. E. Homanics, B. A. Minassian *et al.*, 1998 Mice lacking the beta3 subunit of the GABAA receptor have the epilepsy phenotype and many of the behavioural characteristics of Angelman syndrome. J. Neurosci. **18:** 8505–8514.

Duret, L., D. Mouchiroud and M. Gouy, 1994 Hovergen—a database of homologous vertebrate genes. Nucleic Acids Res. **22:** 2360–2365.

Efstratiadis, A., 1994 Parental imprinting of autosomal mammalian genes. Curr. Opin. Genet. Dev. **4:** 265–280.

Eyre-Walker, A., 1991 An analysis of codon usage in mammals: selection or mutation bias? J. Mol. Evol. **33:** 442–449.

Genetics Computer Group, 1994 Program Manual for the Wisconsin Package, Version 8, September 1994, Genetics Computer Group, 575 Science Drive, Madison, Wisconsin 53711.

Haldane, J. B. S., 1947 The mutation rate of the gene for hemophilia and its segregation ratios in males and females. Ann. Eugen. **13:** 262–271.

Hughes, A. L., and M. Yeager, 1997 Comparative evolutionary rates of introns and exons in murine rodents. J. Mol. Evol. **45:** 125–130.

Hurst, L. D., and H. Ellegren, 1998 Sex biases in the mutation rate. Trends Genet. **14:** 446–452.

Kendrew, J. (Editor), 1994 *The Encyclopedia of Molecular Biology.* Blackwell, Oxford.

Ketterling, R. P., E. Vielhaber, C. D. K. Bottema, D. J. Schaid, M. P. Cohen *et al.*, 1993 Germ-line origins of mutation in families with hemophilia-B: the sex-ratio varies with the type of mutation. Am. J. Hum. Genet. **52:** 152–166.

Kimura, M., 1980 A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. J. Mol. Evol. **16:** 111–120.

Kimura, M., 1983 *The Neutral Theory of Evolution.* Cambridge University Press, Cambridge, United Kingdom.

Li, W. H., 1993 Unbiased estimation of the rates of synonymous and nonsynonymous substitution. J. Mol. Evol. **36:** 96–99.

Li, W. H., 1997 *Molecular Evolution.* Sinauer Associates, Sunderland, MA.

Li, W. H., C. I. Wu and C. C. Luo, 1985 A new method for estimating synonymous and nonsynonymous rates of nucleotide substitution considering the relative likelihood of nucleotide and codon changes. Mol. Biol. Evol. **2:** 150–174.

Makalowski, W., and M. S. Boguski, 1998 Evolutionary parameters of the transcribed mammalian genome: An analysis of 2,820 orthologous rodent and human sequences. Proc. Natl. Acad. Sci. USA **95:** 9407–9412.

McVean, G. T., and L. D. Hurst, 1997 Evidence for a selectively favourable reduction in the mutation rate of the X chromosome. Nature **386:** 388–392.

McVean, G. T., L. D. Hurst and T. Moore, 1996 Genomic evolution in mice and men—imprinted genes have little intronic content. Bioessays **18:** 773–775.

Meguro, M., K. Mitsuya, H. Sui, K. Shigenami, H. Kugoh *et al.*, 1997 Evidence for uniparental, paternal expression of the human GABA(A) receptor subunit genes, using microcell-mediated chromosome transfer. Hum. Mol. Genet. **6:** 2127–2133.

Miller, N., A. H. McCann, D. O'Connell, I. S. Pedersen, V. Spiers *et al.*, 1997 The *MAS* proto-oncogene is imprinted in human breast tissue. Genomics **46:** 509–512.

Miyata, T., H. Hayashida, K. Kuma, K. Mitsuyasu and T. Yasunaga, 1987 Male-driven molecular evolution: a model and nucleotide sequence analysis. Cold Spring Harbor Symp. Quant. Biol. **52:** 863–867.

Monk, M., M. Boubelik and S. Lehnert, 1987 Temporal and regional changes in DNA methylation in the embryonic, extraembryonic and germ cell lineage during mouse embryo development. Development **99:** 371–386.

Moriyama, E. N., and J. R. Powell, 1997 Synonymous substitution rates in Drosophila: Mitochondrial versus nuclear genes. J. Mol. Evol. **45:** 378–391.

Moriyama, E. N., and J. R. Powell, 1998 Gene length and codon usage bias in Drosophila melanogaster, Saccharomyces cerevisiae and Escherichia coli. Nucleic Acids Res. **26:** 3188–3193.

Morton, B. R., and M. T. Clegg, 1995 Neighboring base composition is strongly correlated with base substitution bias in a region of the chloroplast genome. J. Mol. Evol. **41:** 597–603.

Morton, B. R., V. M. Oberholzer and M. T. Clegg, 1997 The

influence of specific neighboring bases on substitution bias in noncoding regions of the plant chloroplast genome. J. Mol. Evol. **45**: 227–231.

MOUCHIROUD, D., C. GAUTIER and G. BERNARDI, 1995 Frequencies of synonymous substitutions in mammals are gene-specific and correlated with frequencies of nonsynonymous substitutions. J. Mol. Evol. **40**: 107–113.

NEUMANN, B., P. KUBICKA and D. P. BARLOW, 1995 Characteristics of imprinted genes. Nat. Genet. **9**: 12–13.

ODANO, I., T. ANEZAKI, H. OHKUBO, Y. YONEKURA, Y. ONISHI *et al.*, 1996 Decrease in benzodiazepine receptor-binding in a patient with Angelman Syndrome detected by I-123 iomazenil and single-photon emission tomography. Eur. J. Nucl. Med. **23**: 598–604.

PLASS, C., H. SHIBATA, I. KALCHEVA, L. MULLINS, N. KOTELEVTSEVA *et al.*, 1996 Identification of Grf1 on mouse chromosome 9 as an imprinted gene by RLGS-M. Nat. Genet. **14**: 106–109.

RICE, P., 1997 *Program Manual for the EGCG Package.* The Sanger Centre, Hinxton Hall, Cambridge, CB10 1RQ, England.

RIESEWIJK, A. M., M. T. SCHEPENS, E. M. MARIMAN, H. H. ROPERS and V. M. KALSCHEUER, 1996 The *MAS* protooncogene is not imprinted in humans. Genomics **35**: 380–382.

SAITOH, S., N. HARADA, Y. JINNO, K. HASHIMOTO, K. IMAIZUMI *et al.*, 1994 Molecular and clinical-study of 61 angelman-syndrome patients. Am. J. Med. Genet. **52**: 158–163.

SCHWEIFER, N., P. J. M. VALK, R. DELWEL, R. COX, F. FRANCIS *et al.*, 1997 Characterization of the C3 YAC contig from proximal mouse chromosome 17 and analysis of allelic expression of genes flanking the imprinted *Igf2r* gene. Genomics **43**: 285–297.

SHARP, P. M., and W. H. LI, 1989 On the rate of DNA sequence evolution in *Drosophila*. J. Mol. Evol. **28**: 398–402.

SHIELDS, D. C., P. M. SHARP, D. G. HIGGINS and F. WRIGHT, 1988 Silent sites in Drosophila genes are not neutral—evidence of selection among synonymous codons. Mol. Biol. Evol. **5**: 704–716.

SHIMMIN, L. C., B. H. J. CHANG, D. HEWETT-EMMETT and W. H. LI, 1993 Potential problems in estimating the male-to-female mutation rate ratio from DNA sequence data. J. Mol. Evol. **37**: 160–166.

SHIMMIN, L. C., B. H. J. CHANG and W. H. LI, 1994 Contrasting rates of nucleotide substitution in the X-linked and Y-linked zinc-finger genes. J. Mol. Evol. **39**: 569–578.

SMITH, N. G. C., and L. D. HURST, 1998a Molecular evolution of an imprinted gene: repeatability of patterns of evolution within the mammalian insulin-like growth factor type II receptor. Genetics **150**: 823–833.

SMITH, N. G. C., and L. D. HURST, 1998b Sensitivity of patterns of molecular evolution to alterations in methodology: a critique of Hughes and Yeager. J. Mol. Evol. **47**: 493–500.

SOKAL, R. R., and F. J. ROHLF, 1995 *Biometry.* W. H. Freeman and Company, New York.

TAMURA, K., 1992 Estimation of the number of nucleotide substitutions when there are strong transition-transversion and G+C content biases. Mol. Biol. Evol. **10**: 512–526.

TERADA, N., K. KUMA and T. MIYATA, 1997 Verification of male-driven molecular evolution. Genes Genet. Syst. **72**: 428.

THOMPSON, J. D., D. G. HIGGINS and T. J. GIBSON, 1994 ClustalW—improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. **22**: 4673–4680.

VILLAR, A. J., and R. A. PEDERSEN, 1994 1994 parental imprinting of the *Mas* protooncogene in mouse. Nat. Genet. **8**: 373–379.

WENTWORTH, B. M., I. M. SCHAEFER, L. VILLA-KOMAROFF and J. M. CHIRGWIN, 1986 Characterization of the two nonallelic genes encoding mouse preproinsulin. J. Mol. Evol. **23**: 305–312.

WOLFE, K. H., and P. M. SHARP, 1993 Mammalian gene evolution—nucleotide-sequence divergence between mouse and rat. J. Mol. Evol. **37**: 441–456.

WRIGHT, F., 1990 The effective number of codons used in a gene. Gene **87**: 23–29.

YANG, Z., 1997 PAML: a program package for phylogenetic analysis by maximum likelihood. Comput. Appl. Biosci. **13**: 555–556.

YANG, W., K. CHEN, N. C. LAN, T. K. GALLAHER and J. C. SHIH, 1992 Gene structure and expression of the mouse 5-HT2 receptor. J. Neurosci. Res. **33**: 196–204.

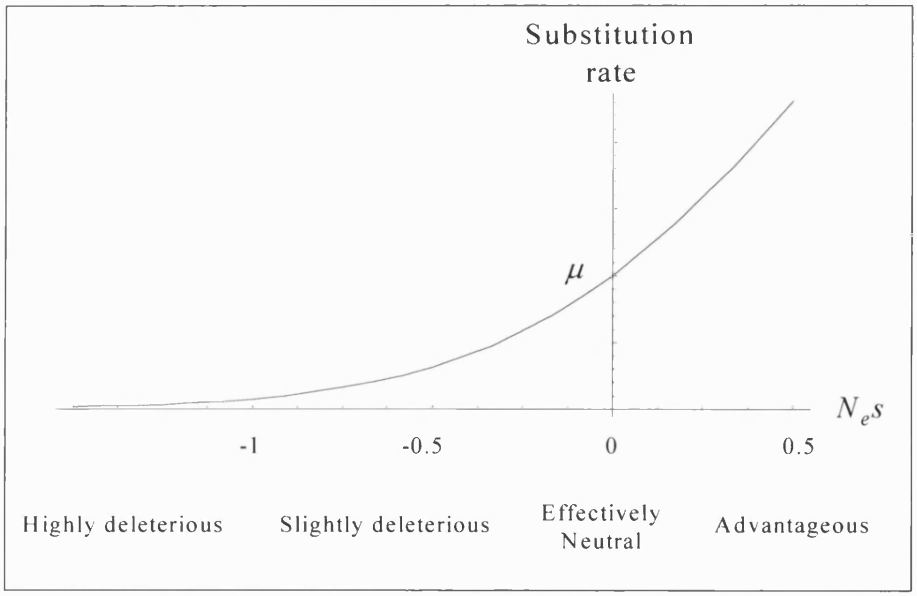
Communicating editor: R. R. HUDSON

Chapter 6. Why do genes evolve at different rates?

Evolutionary rates fit into the paradigm of genomic anatomy because the rate of evolution is a measure of the speed at which a gene's identity changes, and identity is one of the three factors of genomic anatomy (see Chapter 1). So why do genes evolve at different rates (as shown by Wolfe and Sharp 1993)? The answer is that the processes of mutation, selection and random drift must also vary between genes: mutation need not be constant across the genome, fitness is unlikely to affected equally by mutations at all sites, and the strength of random drift will vary across the genome as a result of recombination rate variation and Hill-Robertson (1966) effects. Later on I shall describe a test of one explanation for variation in rates of protein evolution, but first I shall describe some of the underlying theory.

The relative strengths of the major evolutionary processes will depend on the type of site at which evolution is measured. If mutations at freely recombining synonymous sites in mammals are neutral, that is unaffected by selection, then K_S will reflect the underlying mutation rate (Kimura 1983). Then the ratio of K_A/K_S enables a measure of protein evolution controlling for variation in the mutation rate. Thus K_A/K_S should provide an indication of the nature of selection: if mutations are deleterious then K_A/K_S decreases below one and if mutations are advantageous then K_A/K_S increases above one.

The interplay between selection and random drift depends on the product of the coefficient of selection (s) and the effective population size (N_e). When selection is strong ($N_e s >> 1$) then deleterious mutations are substituted at a very low rate relative to neutral mutations, and advantageous mutations are substituted at a very high rate relative to neutral mutations. When selection is weak ($N_e s < 1$) then the substitution rate differences with respect to neutral mutations are less extreme for both advantageous and deleterious mutations. These effects are illustrated in the figure below, a plot of equation 3.14 from Kimura's book (1983).



Alternative theories of molecular evolution differ with respect to the influence of selection on substitution rates. Under the neutral theory only neutral mutations contribute to substitution rates, since advantageous mutations are very rare and selection against deleterious mutations is strong. Under the nearly neutral theory both neutral and slightly deleterious mutations contribute to the substitution rate. Under selectionist theories substitution rates are determined by slightly deleterious, effectively neutral and advantageous mutations. So in order to explain variation in molecular evolutionary rates we need to determine the variation in two factors: (1) the sensitivity of organismal fitness to gene mutation and (2) the ratio of effectively deleterious to effectively advantageous mutations.

The sensitivity of organismal fitness to gene mutation is dependent on the two parameters identified by Wilson *et al.* (1977): P is the sensitivity of gene function to mutations and Q is the sensitivity of organismal fitness to gene function. The concept of P underpins two of Kimura and Ohta's principles of molecular evolution (as given in Kimura 1983): "Functionally less important molecules or parts of molecules evolve (in terms of mutant substitutions) faster than more important ones" and "Those mutant substitutions that are less disruptive to the existing structure and function of a molecule (conservative substitutions) occur more frequently in evolution than more disruptive ones".

P is likely to depend on a number of factors (see section 9.5 in Kimura 1983). A smaller proportion of neutral sites is expected in multimers versus monomers owing to their intersubunit interactions. The larger the molecule the greater the proportion of amino acids performing the weakly selected function of space filling. A lower proportion of neutral mutations is expected for substrate specific genes.

Q is also likely to vary between genes. Ubiquitous housekeeping genes may be subjected to different selection pressures from those affecting tissue specific genes. Highly abundant insoluble proteins are likely to be subject to stronger selection than less abundant soluble enzymes, hence fewer mutations will be neutral.

If there are multiple gene copies per haploid genome, or if gene functions are to some extent redundant, or if the mutations are recessive then changes to gene function are more likely to be neutral. However, this logic may not apply for the same reason that diploidy leads to a higher mutation load than haploidy: if selection is weak then mutations will accumulate in the population and so gene function may be affected by multiple mutations. The effect of additional mutations is then indicated by epistasis and dominance effects. A further complication is that effective copy number itself might be under selection, for example duplications of particularly important developmental genes might be selectively favoured as a guard against failures in the control of gene expression.

One difficulty with testing the prediction that P and Q should affect rates of evolution is in obtaining reliable measures of these parameters (Wilson *et al.* 1977). However, knockout

phenotype appears to provide a direct measure of Q, the dispensability of a protein (see Research Paper 7).

The ratio of effectively deleterious to effectively advantageous mutations seems harder to predict. Note that the problem here is different to the usual form of the neutralist-selectionist debate. The question is not whether advantageous mutations exist, but whether genes vary in their relative proportions of advantageous and deleterious mutations. Although this ratio could hardly vary if there were no advantageous mutations, the neutralist position is that advantageous mutations do occur, but so rarely that such mutations have little effect on substitution rates. Even if advantageous mutations are rare in most genes, they may be common in some genes. One factor that will clearly influence the proportion of advantageous mutations is the degree to which the required function of the gene will change over time. For example, the required function of a gene which codes for an unspecific DNA binding protein will not change since the structure of DNA will not change, but the required function of a parasite antigen receptor will swiftly change as the parasite evolves new antigens. Such reasoning may well explain why histones have such a low K_A/K_S ratio and immunoglobulins such a high K_A/K_S ratio.

One can predict that ongoing evolutionary races are likely to generate high ratios of advantageous to deleterious mutations. The problem comes in predicting whether evolutionary races will persist or be resolved. For example, it is unclear whether the conflict theory of imprinting predicts ongoing conflict or an ESS (see Chapter 5).

How can we test explanations for varying substitution rates? As mentioned above, the knockout phenotype provides a direct measure of Q. So under the nearly neutralist theory (mutations are strongly deleterious, slightly deleterious or neutral) we obtain the prediction that genes with extreme knockout phenotypes should have lower K_A/K_S ratios than genes with mild knockout phenotypes.

I have tested this prediction using a set of mouse-rat orthologs, divided into two groups on the basis of knockout phenotype (see Research Paper 7) .The results of sequence analysis initially appear to support the hypothesis: K_A/K_S is significantly higher for the non-essential genes.

However the two groups of genes may not be equal with respect to those other factors, discussed above, which may also affect evolutionary rates. In particular the non-essential genes contain a much higher proportion of immune system genes than the essential genes. The assumption of no advantageous mutations (in nearly neutral theory) is particularly unlikely to apply to immune system genes (Hughes and Nei 1988), and thus immune system genes would be expected to show high K_A/K_S ratios, as this analysis confirms. If immune system genes are excluded from the analysis, then essential and non-essential genes do not evolve at different rates. The differences between essential and non-essential genes can also be examined using neuronal genes as a control for tissue specificity which may affect evolutionary rates. Neuronal essential genes do not differ evolve at different rates from neuronal non-essential genes. Thus knockout phenotype does not explain variation in molecular evolutionary rates.

The effect of tissue specificity can be further examined by comparing neuronal genes with reproductive genes. Neuronal genes have low K_A/K_S ratios but reproductive genes do not. One factor which might complicate the interpretation of these comparisons is the possibility that reproductive genes may be involved in ongoing evolutionary races due to conflict between mother and offspring (Wallis 1981; Wallis 1993) or sexual conflict (Tsaur and Wu 1997). Thus reproductive genes may be affected by a higher proportion of advantageous mutations than neuronal genes, and so differences in evolutionary rates between these two classes of genes may not be solely due to the strength of selection on deleterious mutations.

Why does knockout phenotype fail to covary with evolutionary rates? One explanation is that knockout phenotype, in the absence of wildtype environments, tells us little about the effect of a knockout in the wild. Another explanation is that the effect of knockouts may not reflect the effect of substitutional changes. Although Gillespie (1991) has argued that you can hardly get a greater selective effect than from a null mutation, some substitutional mutations might cause gains of function which could not appear with knockouts. A final explanation is that perhaps selection is very powerful. Thus a difference in the selective coefficient makes little difference to substitution rates if selection is strong ($N_e s >> 1$). This is equivalent to noting that the slope of substitution rate against $N_e s$ is nearly flat when $N_e s$ is large and negative (see Figure above).

Research Paper 7. Do essential genes evolve slowly?

Laurence Hurst and Nick Smith (1999)
Current Biology **9** 747-750.

Do essential genes evolve slowly?

Laurence D. Hurst and Nick G.C. Smith

Approximately two thirds of all knockouts of individual mouse genes give rise to viable fertile mice. These genes have thus been termed 'non-essential' in contrast to 'essential' genes, the knockouts of which result in death or infertility. Although non-essential genes are likely to be under selection that favours sequence conservation [1], it is predicted that they are less subject to such stabilising selection than essential genes, and hence evolve faster [2]. We have addressed this issue by analysing the molecular evolution of 108 non-essential and 67 essential genes that have been sequenced in both mouse and rat. On preliminary analysis, the non-essential genes appeared to be faster evolving than the essential ones. We found, however, that the non-essential class contains a disproportionate number of immune-system genes that may be under directional selection (that is, selection favouring change) because of host–parasite coevolution. After correction for this bias, we found that the rate at which genes evolve does not correlate with the severity of the knockout phenotype. This was corroborated by the finding that, whereas neuron-specific genes have significantly lower rates of change than other genes, essential and non-essential neuronal genes have comparable rates of evolution. Our findings most probably reflect strong selection acting against even very subtle deleterious phenotypes, and indicate that the putative involvement of directional selection in host–parasite coevolution and gene expression within the nervous system explains much more of the variance in rates of gene evolution than does the knockout phenotype.

Address: Department of Biology and Biochemistry, University of Bath, Claverton Down, Bath BA2 7AY, UK.
E-mail: l.d.hurst@bath.ac.uk

Received: **9 April 1999**
Revised: **10 May 1999**
Accepted: **4 June 1999**

Published: **5 July 1999**

Current Biology 1999, 9:747–750
http://biomednet.com/elecref/0960982200900747

 ISSN 0960-9822

Results and discussion

Wilson *et al.* [2] proposed in 1977 that two proteins subject to the same level of functional constraint, but differing in their dispensability, will evolve at different rates. They argued that this may, in part, result from the fact that mutations in redundant genes may be masked. Thus, essential genes (knockouts of which are lethal or infertile) should evolve slower than non-essential genes (knockouts of which are viable and fertile) — this is the 'knockout-rate' prediction. From the adaptive theory of mutation rates, it might similarly be predicted that the mutation rates of essential genes should be lower than those of non-essential genes [3–5]. We have examined these issues using sequence comparison of mouse and rat genes with cross-reference to knockout data (see Materials and methods for details). For any aligned pair of orthologous genes we can calculate K_A, the rate of DNA substitutions affecting the amino-acid composition of the gene product (non-synonymous substitutions), and K_S, the rate of DNA substitutions that are silent at the amino-acid level (synonymous substitutions). If synonymous mutations in mammals are neutral, as may be the case [4], then K_S is a measure of the mutation rate [6] and K_A/K_S is a measure of the rate of protein evolution after controlling for mutation rate.

Rates of evolution and the knockout phenotype

Essential genes ($n = 67$ in our sample) had significantly lower K_A/K_S ratios than non-essential genes ($n = 108$) (Mann-Whitney U-test, $p = 0.0009$, Figure 1; all results and statistics are given in Table 1). Essential genes also tended to have lower K_S values than non-essential genes ($p = 0.04$ in a one-tailed test).

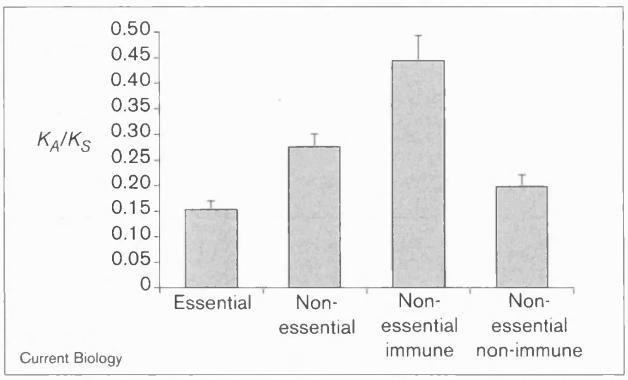
Rates of evolution after controlling for 'tissue of activity'

The above results are potentially confounded by differences between essential and non-essential genes in the types of tissues that they act in. Notably, the non-essential class contained many more genes specific to the immune system than did the essential class (34/108 versus 3/67, $p < 10^{-5}$ using Fisher's exact test). Genes of the immune system tend to have high K_A/K_S ratios [7,8] and high K_S values [7]. We confirmed these results: K_A/K_S ratios of immune-system genes ($n = 37$) were on average more than double those of non-immune genes ($n = 138$, $p = 0.0028$) and immune-system genes also had higher K_S values ($p = 0.02$) (Table 1). The high K_A/K_S ratios can be accounted for by arguing that immune-system genes are, at least along part of their sequence, likely to be under directional selection (and/or overdominance) driven by host–parasite coevolution (see for example [9–11]).

We then asked whether the non-essential non-immune genes ($n = 74$) had higher K_A/K_S and higher K_S values than essential non-immune genes ($n = 64$). When we did this comparison we found no significant differences ($p = 0.24$ and $p = 0.45$, respectively; Table 1).

The effect of tissue specificity on rates of gene evolution can also be controlled for by examining genes expressed specifically in neurons. Neither the K_A/K_S ratios nor the

Figure 1



The K_A/K_S ratios (± SEM) for various classes of gene. The left-hand two columns give the values for essential genes ($n = 67$) and non-essential genes ($n = 108$), respectively. These are significantly different ($p < 0.001$). The right-hand two columns show the data for the non-essential genes broken down into those with immune function ($n = 34$) and those without ($n = 74$). Note that the major difference between the essential and non-essential genes is largely due to the excess of immune-related genes in the latter class.

K_S values for non-essential neuronal genes ($n = 18$) were significantly different from those of the essential neuronal genes ($n = 16$; $p = 0.63$ and $p = 0.36$, respectively). The comparison between all neuronal genes ($n = 34$) and non-immune non-neuronal genes ($n = 104$) revealed that K_A/K_S values for neuronal genes were less than half those of the others ($p = 0.0001$). The two classes had similar K_S values ($p = 0.51$).

Is this low rate of evolution a peculiarity of neuronal genes or a general property of tissue-specific genes? This is hard to evaluate, but the fact that genes specifically involved in reproduction ($n = 18$) had higher K_A/K_S ratios than neuronal genes ($n = 34$, $p = 0.013$) suggests that the low rate is peculiar to neuronal genes. This was further corroborated by the finding that reproductive genes had similar K_A/K_S ratios and K_S values to those of the 'normal' class of non-immune non-neuronal genes ($p = 0.91$ and $p = 0.31$, respectively).

The effect of tandem substitutions

In the mouse–rat comparison there is a strong correlation between K_A and K_S (see for example [12,13]). It seems that this correlation is in large part due to an excess of tandem (adjacent) substitutions, which may be the result of selection (N.G.C.S. and L.D.H., unpublished observations). If selection acting on synonymous mutations is responsible for the excess of tandem substitutions, then K_S is not an unbiased measure of the mutation rate. We therefore re-analysed our sequences ignoring tandem substitutions.

Our results concerning stabilizing selection were not affected when tandem substitutions were ignored. The non-essential genes had a significantly higher K_A/K_S ratio than the essential genes ($p = 0.0007$). This result was again attributable to the excess of immune-system genes within the non-essential class, as the K_A/K_S ratios are similar for non-immune essential and non-immune non-essential genes ($p = 0.288$). Immune-system genes had high K_A/K_S ratios compared with all non-immune genes ($p < 0.0001$). Neuronal genes had especially low K_A/K_S ratios compared with non-immune non-neuronal genes ($p < 0.0001$). In contrast, our results as regards K_S were altered when tandem substitutions were ignored. The essential genes had mutation rates no lower than those of non-essential genes ($p = 0.25$) and the immune genes had K_S values no higher than non-immune genes ($p = 0.44$). Thus, the variation in 'raw' K_S values might partly have been due to selection on synonymous mutations. With the effects of such selection removed, we found no evidence of adaptive variation in mutation rates. Hence, we need only seek to explain the variation in K_A/K_S ratios.

Why is there no correlation between knockout phenotype and rates of evolution?

Our results suggest that knockout phenotype does not correlate with the K_A/K_S ratio once we have accounted for tissue-specific effects. To explain this result it is worth clarifying the logic underlying the knockout-rate prediction. The knockout phenotype indicates the magnitude of the selective difference between the presence and absence of the gene: the more severe the knockout phenotype, the greater the strength of selection on the entire gene. The knockout-rate prediction requires not only that the strength of selection on the entire gene should co-vary with the average strength of selection on non-synonymous mutations within the gene (although each mutation is likely to be under weaker selection than the null mutation) but, crucially, that the strength of selection (s) on these individual mutations (the magnitude of their effects on fitness) is within certain bounds.

Imagine that non-synonymous mutations in non-essential genes tend to have, on average, a smaller effect on fitness than do mutations in essential genes. This will not lead to differences in the rate of fixation of mutations if the mutations are highly deleterious: in a reasonably sized population a mutation reducing fitness by 10% will be no more likely to reach fixation than one that reduces fitness by 20%. But for very low s in a finite population, however, deleterious mutations can reach fixation (become a substitution) by random drift. In this case, the size of s affects the fixation probability and hence the substitution rate. So one way of arriving at the knockout-rate prediction is to suppose that non-essential genes have more non-synonymous mutations with s values that are compatible with fixation. Our failure to find any differences in rates of evolution between essential and non-essential genes might then be due to the fact that most non-synonymous mutations in the two types of genes are highly deleterious

Table 1

The *p* values from two-tailed Mann-Whitney U-tests in pairwise comparisons of various gene classes.

Class	*n*	K_A/K_S	NE	E	I-NE	I	NI-E	NI-NE	NR	NNR-NI	NR-E	NR-NE	NI	R
		K_S	0.186 ± 0.007	0.165 ± 0.007	0.209 ± 0.014	0.204 ± 0.013	0.165 ± 0.008	0.177 ± 0.008	0.163 ± 0.011	0.174 ± 0.006	0.148 ± 0.014	0.176 ± 0.017	0.171 ± 0.005	0.152 ± 0.016
Non-essential (NE)	108	0.275 ± 0.024	–	0.0763	*0.1495*	*0.2411*	0.0904	*0.3878*	*0.1384*	*0.2410*	0.0838	*0.5890*	*0.1321*	*0.1143*
Essential (E)	67	0.153 ± 0.017	**0.0009**	–	**0.0062**	***0.0114***	*0.9725*	0.4172	*0.9227*	*0.4854*	*0.4706*	0.6019	*0.6042*	*0.5944*
Immune non-essential (I-NE)	34	0.444 ± 0.048	***0.0005***	**0.0000**	–	*0.7779*	**0.0080**	**0.0461**	**0.0170**	**0.0226**	**0.0115**	0.1632	**0.0116**	**0.0205**
Immune (I)	37	0.418 ± 0.047	***0.0023***	***0.0000***	*0.6166*	–	**0.0143**	0.0762	**0.0275**	**0.0399**	**0.0176**	0.2192	**0.0212**	**0.0307**
Non-immune essential (NI-E)	64	0.155 ± 0.018	**0.0012**	*0.9962*	**0.0000**	**0.0000**	–	0.4460	*0.9020*	*0.5189*	*0.4595*	0.6181	*0.6405*	*0.5792*
Non-immune non-essential (NI-NE)	74	0.198 ± 0.022	***0.0356***	0.2301	**0.0000**	**0.0000**	0.2445	–	*0.4468*	*0.8560*	0.2457	*0.9647*	*0.6776*	*0.2991*
Neuronal (NR)	34	0.096 ± 0.02	***0.0000***	***0.0196***	**0.0000**	**0.0000**	***0.0220***	*0.0014*	–	0.5127	*0.5815*	*0.6103*	*0.6104*	0.7583
Non-neuronal non-immune (NNR-NI)	104	0.20 ± 0.017	*0.0661*	*0.0610*	**0.0000**	**0.0000**	*0.0694*	*0.5963*	**0.0001**	–	0.2613	0.9281	*0.8059*	*0.3104*
Neuronal essential (NR-E)	16	0.081 ± 0.019	**0.0005**	*0.0934*	**0.0000**	**0.0000**	*0.1006*	**0.0145**	0.7640	**0.0037**	–	0.3605	0.3070	0.8360
Neuronal non-essential (NR-NE)	18	0.109 ± 0.038	***0.0005***	0.0550	**0.0000**	**0.0000**	*0.0578*	***0.0141***	*0.7902*	**0.0040**	0.6255	–	*0.8266*	0.4765
Non-immune (NI)	138	0.177 ± 0.015	***0.0013***	*0.4894*	**0.0000**	**0.0028**	*0.5066*	*0.4887*	***0.0027***	*0.1552*	***0.0313***	***0.0211***	–	*0.3838*
Reproductive (R)	18	0.193 ± 0.037	*0.2676*	*0.2781*	**0.0009**	**0.0028**	*0.2882*	*0.8982*	**0.0130**	*0.9138*	**0.0329**	**0.0354**	*0.5371*	–

In the lower left half of the table are the *p* values for each of the pairwise K_A/K_S comparisons between all of the gene classes and in the upper right half are the *p* values for the K_S comparisons. The sample size (*n*) and the mean K_A/K_S and K_S values (± SEM) for each gene class are given. Significant *p* values are shown in bold. Those shown in italics are comparisons in which there is overlap between data sets. The genes in this analysis have not had tandem substitutions removed. A significant statistic indicates that the set of genes in one class is significantly greater or smaller in the parameter in question than the other set of genes. To establish which is the greater, it is necessary to consult the absolute mean values for K_A/K_S or K_S given in column 3 and row 2, respectively.

(that is, selection is very strong), and that the proportion of slightly deleterious or effectively neutral mutations in the two is about the same.

There may, however, be other explanations. First, there may be methodological problems. The knockout phenotype may be a poor indicator of the fitness effect of a knockout in nature, owing to the artificiality of the laboratory environment. Alternatively, the mouse–rat data may be misleading. Mutational saturation of sites is a potential problem in any analysis of this sort. A sizeable majority of silent sites in the mouse–rat comparison have not changed, however, indicating that saturation is unlikely. Second, some of the knockout-rate prediction's other assumptions may be invalid. The prediction also requires, for example, that advantageous mutations are rare. Analysis of non-synonymous evolution in immune-system genes and in hemizygously expressed genes [4,5] indicates that this may be unrealistic. How this might affect the predictions is unclear.

Why is there a correlation between tissue specificity and rates of evolution?

We have found that the rate of evolution as measured by the K_A/K_S ratio is strongly affected by tissue specificity: immune-system genes evolve rapidly, neuronal genes evolve slowly and reproductive genes evolve at an intermediate rate. There are a number of possible explanations for an effect of tissue specificity on K_A/K_S values. Genes in different tissues might by affected by different proportions

of advantageous and deleterious mutations. A high proportion of advantageous mutations would seem to explain the high rates of evolution of immune-system genes. Tissues may also differ in their influence on an organism's fitness, although this is, as yet, mere speculation. But we can be sure that the putative involvement in host–parasite coevolution and specific expression within the nervous system explain the variance in the rate of protein evolution much better than do knockout phenotypes.

Materials and methods

Database assembly

The gene knockout database (gkd: www.bioscience.org/knockout/knochome.htm) annotates the phenotypes of over 300 single-gene knockouts in mice. We also used four genes annotated in Tbase (tbase.jax.org/) but not present in gkd. Knockouts reported as infertile or inviable were incorporated into the class of essential genes. Genes whose knockout had sex-specific infertility or high, but not complete, levels of inviability were also included within this class. Where some phenotype was observed (in 97% of cases) the genes are classified in the gkd by the tissue principally disrupted. Some are classed as multi-tissue if no one organ or system is predominantly affected. We checked all genes against the relevant Tbase entries or original papers to confirm the designation of the 'tissue of activity' and the viability/fertility. Internet links to the genome database (gdbwww.gdb.org/gdb/), mouse genome informatics (mgd.hgmp.mrc.ac.uk/), Online Mendelian Inheritance in Man (www3.ncbi.nlm.nih.gov/omim/) or to the original paper, were followed to obtain links to the GenBank citation for the gene concerned.

For each gene the HOVERGEN database of vertebrate homologous genes [14] was searched using the GenBank entry name. From observation of the phylogeny of each gene family, we assembled a list of accession numbers for the mouse and rat orthologues corresponding to known gene knockouts. The accession numbers of the 108 non-essential and 69 essential genes are provided in the Supplementary material.

Preparation of alignments

FETCH was used to extract sequences from databases and GENETRANS was used to extract and combine exons automatically. Protein alignments were performed using CLUSTALW [15]. The DNA alignments were re-created from the protein alignments and the original DNA sequences using the program MRTRANS (written by W. Pearson, and available at HGMP (www.hgmp.mrc.ac.uk/)). Alignments were performed using the GCG [16] and EGCG [17] packages at HGMP. On close inspection two of the essential genes were found to have dubious alignments and were eliminated, reducing the sample size to 67. For these and the 108 non-essential genes we calculated the number of gaps in the alignments, in order to ask whether the alignment protocol might be affecting our rate estimates. Using both the total number of gaps and the number of gaps per informative site, we found not even a weak correspondence with any of the rate parameters. Alignment artefacts are probably not producing systematically biased rate estimates.

Algorithmic rate estimation

Substitution rates were estimated from alignments using methods developed by Moriyama and Powell [18]. Tamura and Nei's [19] multiple hits correction method was used to account for variation in base composition, and Li's [20] method was used to calculate the non-synonymous rate per site (K_A) and the synonymous rate per site (K_S).

Supplementary material

Supplementary material including a complete list of the genes analysed is available at http://current-biology.com/supmat/supmatin.htm.

Acknowledgements

We wish to thank Mike Danson and Adam Eyre-Walker for discussion.

References

1. Brookfield JFY: **Genetic redundancy: screening for selection in yeast.** *Curr Biol* 1997, **7**:R366-R368.
2. Wilson AC, Carlson SS, White TJ: **Biochemical evolution.** *Annu Rev Biochem* 1977, **46**:573-639.
3. Cox EC: **On the organization of higher chromosomes.** *Nat New Biol* 1972, **92**:133-134.
4. McVean GT, Hurst LD: **Evidence for a selectively favourable reduction in the mutation rate of the X chromosome.** *Nature* 1997, **386**:388-392.
5. Smith GC, Hurst LD: **The causes of synonymous rate variation in the rodent genome: can substitution rates be used to estimate the sex bias in mutation rate?** *Genetics* 1999, in press.
6. Kimura M: *The Neutral Theory of Evolution.* Cambridge, UK: Cambridge University Press; 1983.
7. McVean GT, Hurst LD: **Molecular evolution of imprinted genes: no evidence for antagonistic coevolution.** *Proc R Soc Lond B* 1997, **264**:739-746.
8. Kuma K, Iwabe N, Miyata T: **Functional constraints against variations on molecules from the tissue level – slowly evolving brain-specific genes demonstrated by protein-kinase and immunoglobulin supergene families.** *Mol Biol Evol* 1995, **12**:123-130.
9. Hughes AL: **Rapid evolution of immunoglobulin superfamily C2 domains expressed in immune system cells.** *Mol Biol Evol* 1997, **14**:1-5.
10. Hughes AL, Nei M: **Pattern of nucleotide substitution at major histocompatibility complex class I loci: evidence for overdominant selection.** *Nature* 1988, **335**:167-170.
11. Hughes AL, Ota T, Nei M: **Positive Darwinian selection promotes charge profile diversity in the antigen binding cleft of class I MHC molecules.** *Mol Biol Evol* 1990, **7**:515-524.
12. Wolfe KH, Sharp PM: **Mammalian gene evolution: nucleotide sequence divergence between mouse and rat.** *J Mol Evol* 1993, **37**:441-456.
13. Makalowski W, Boguski MS: **Synonymous and nonsynonymous substitution distances are correlated in mouse and rat genes.** *J Mol Evol* 1998, **47**:119-121.
14. Duret L, Mouchiroud D, Gouy M: **Hovergen – a database of homologous vertebrate genes.** *Nucl Acid Res* 1994, **22**:2360-2365.
15. Thompson JD, Higgins DG, Gibson TJ: **Clustal-W – improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice.** *Nucleic Acid Res* 1994, **22**:4673-4680.
16. GCG: *Program Manual for the Wisconsin Package,Version 8.* Wisconsin, USA: Genetics Computer Group; 1994.
17. Rice P: *Program Manual for the EGCG Package.* Cambridge, UK: The Sanger Centre; 1997.
18. Moriyama EN, Powell JR: **Synonymous substitution rates in *Drosophila*: mitochondrial versus nuclear genes.** *J Mol Evol* 1997, **45**:378-391.
19. Tamura K, Nei M: **Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees.** *Mol Biol Evol* 1993, **10**:512-526.
20. Li W-H: **Unbiased estimation of the rates of synonymous and nonsynonymous substitution.** *J Mol Evol* 1993, **36**:96-99.

Chapter 7. Evolution at synonymous sites

Synonymous mutations in protein-coding genes affect DNA and mRNA sequences but not amino acid sequences, and occur as a consequence of the redundancy of the genetic code. Since the phenotype is primarily determined at the protein level, it might be thought that synonymous mutations which are silent at the protein level should be unaffected by selection. When the neutral theory of molecular evolution was first proposed (Kimura 1968a; King and Jukes 1969), the neutrality of synonymous mutations was one of its most important tenets (Powell and Moriyama 1997), and was consistent with what was known at the time: "As far as is known, synonymous mutations are truly neutral with respect to natural selection" (King and Jukes 1969). However, the neutral theory does not suggest that all mutations are neutral (Kimura 1983), and Kimura (1968b), rather presciently, did not suppose that all synonymous changes would be neutral: "It is important to note that probably not all synonymous mutations are neutral, even if most of them are nearly so".

In the thirty years since the neutral theory was first proposed, it has become clear that selection does act on synonymous mutations, at least in some organisms (Akashi and Eyre-Walker 1998; Hey 1999). Critical attention is now focused on determining how selection affects silent sites and the minimum effective population size required for such selection pressures to affect molecular evolution.

In the context of genomic anatomy, selection at synonymous sites provides a way to make sense of identity at the smallest possible scale of genomic anatomy, at the level of the individual nucleotide. If selection really does act on synonymous mutations, then the patterns of nucleotide identity at silent sites can make sense. I shall briefly consider some general tests of selection on silent sites, before describing more detailed tests of the nature of selection on synonymous sites. Then I shall explain why the question of whether selection acts on silent sites in mammals is an interesting issue, and then describe some of the tests I have performed to examine this question.

Tests of selection on silent sites

Comparisons of substitution rates can help to indicate the processes underlying nucleotide substitution (Li 1997). If one assumes strict neutrality and that mutation rates are uniform within genes, then variation in substitution rates for different classes of site within a gene corresponds to variation in levels of functional constraint (Kimura and Ohta 1974). Under these assumptions the finding that fourfold degenerate sites evolve faster than twofold degenerate sites which in turn evolve faster than nondegenerate sites implies that functional constraints decrease as degeneracy increases (Li and Graur 1991). However, it cannot be deduced that synonymous sites are devoid of any functional constraints. Synonymous rates vary across the genome, although nonsynonymous rates vary to a much greater extent (Wolfe and Sharp 1993). However, in the absence of reliable

data on mutation rate variation across the genome, it is unclear whether such synonymous rate variation is due to selection or mutation.

The existence of a molecular clock, a constant rate of evolution in all evolutionary lineages for a given macromolecule, was first suggested by Zuckerkandl and Pauling (1965). If one assumes a molecular clock and that mutation is a Poisson process then the index of dispersion of substitution rates should be unity (Zuckerkandl and Pauling 1965). The index of dispersion for synonymous substitutions as well as nonsynonymous substitutions is often found to be greater than one (Gillespie 1991; Li 1997), and the existence of any sort of predictive molecular clock has been questioned (Gibbs *et al.* 1998). The overdispersion of synonymous substitutions can be explained by all three classes of molecular evolution theories. Under strict neutrality, the overdispersion can be explained by lineage specific mutation rates, generation time effects (Ohta 1993), or non-independence of mutation events. Under the nearly neutral theory, effective population size changes can explain variation in substitution rates (Ohta 1972). Under selectionist models, changes in selection pressures may cause the variation in rates. This plethora of arguments against a clock may mean that the only reason for an approximate clock comes from the expected convergence of large numbers (Ayala 1999).

Polymorphism data can also be used to test for the presence of selection. The tests of Tajima (1989) and Fu and Li (1993) examine the proposition that all mutations in a DNA region are selectively neutral (Li 1997), and so cannot be applied exclusively to synonymous sites which are interspersed with nonsynonymous sites. The HKA test (Hudson *et al.* 1987) uses the neutral theory prediction of a positive correlation between intra- and inter- population DNA variation, but observations of high silent site within-species polymorphism at the *Adh* locus in *Drosophila melanogaster* (Kreitman and Hudson 1991) could well be due to selection on nonsynonymous rather than synonymous sites. The McDonald-Kreitman test (McDonald and Kreitman 1991) compares the ratio of fixed and polymorphic differences between different classes of sites. Comparisons of nonsynonymous and synonymous substitutions reveal an excess of synonymous polymorphisms which is consistent with selection affecting nonsynonymous sites more than synonymous sites. However the McDonald-Kreitman test can be applied to classes of sites other than synonymous and nonsynonymous (e.g. Jenkins *et al.* 1995), and indeed Akashi (1995) has applied the McDonald-Kreitman test to compare preferred (to the major codon, see below) and unpreferred (from the major codon) synonymous substitutions, thereby inferring weak selection on silent sites in *Drosophila*.

Major codon preference and translational selection

To find conclusive evidence of selection we have to consider in detail both the patterns of silent site identity and the biochemical bases of selection.

Specific patterns of silent site identity have been quantified as codon usage bias, the unequal usage of codons encoding the same amino acid. Codon usage bias can be measured in a

number of ways (Powell and Moriyama 1997; Sharp *et al.* 1986; Shields *et al.* 1988; Wright 1990), and although the alternative measures do differ with respect to the effect of gene length (Comeron and Aguade 1998; Moriyama and Powell 1998), most measures of codon bias are highly correlated (Powell 1997).

Selection is thought to act on silent sites to enhance protein synthesis (Sharp *et al.* 1993). Silent site identity is biased towards "major" codons, which are recognised by abundant tRNAs and which show perfect Watson-Crick pairing with the tRNA anticodon. Such relationships between tRNAs and codon usage have been found in *E. coli* (Ikemura 1985), *S. cerevisiae* (Ikemura 1985), and *D. melanogaster* (Moriyama and Powell 1997). Codon usage bias tends to be higher in highly expressed genes on which selection is thought to be relatively strong, both in yeast and *E. coli* (Ikemura 1985).

The selection-mutation-drift theory of synonymous codon usage suggests that "codon usage patterns result from the balance in a finite population between selection favouring an optimal codon for each amino acid and mutation together with drift allowing the persistence of nonoptimal codons" (Bulmer 1991). Since linkage restricts the efficacy of selection (Barton 1995), codon usage bias should correlate positively with recombination rates, as has been found in *D. melanogaster* (Kliman and Hey 1993). As described above, levels of divergence and polymorphism also indicate selection favouring major codons (Akashi 1995). Silent site divergence has been found to be negatively correlated with codon usage bias in *E. coli* (Sharp and Li 1987) and *D. melanogaster* (Sharp and Li 1989), although such relationships may be partially due to the relationship between silent site divergence and expression levels (Akashi and Eyre-Walker 1998).

The biochemical basis of major codon preference is thought to stem from either improved speed or accuracy of translation. The relationship between tRNA pools and codon use is consistent with selection for translation elongation rates (Berg and Kurland 1997). Higher codon usage in amino acid conserved positions in *D. melanogaster* is consistent with selection for translational accuracy (Akashi 1994), although such a relationship is not found in *E. coli* (Hartl *et al.* 1994), and may be a consequence of the correlation between silent and replacement divergence (Akashi and Eyre-Walker 1998) which has many alternative explanations (see Li 1997). Selection for translational accuracy is consistent with the correlation between codon bias and gene length found in *E. coli* (Eyre-Walker 1996) and yeast (Moriyama and Powell 1998), although the relationship between codon bias and gene length is complicated by correlated gene expression and linkage effects (Akashi and Eyre-Walker 1998).

Translational selection need not always act to maximise translational speed. Programmed frameshift events in *E. coli* depend on the use of minor codons (Schwartz and Curran 1997), and proper protein folding may depend on translational pauses (Purvis *et al.* 1987).

Alternative forms of selection are suggested by the reduced codon biases and silent divergence rates found at the start of *E. coli* (Eyre-Walker and Bulmer 1993) and *D. melanogaster*

(Kliman and Eyre-Walker 1998) genes. Minor codons are also found in the middle of genes from a wide variety of enteric bacteria, and do not appear to be the result of site specific mutation biases (Maynard Smith and Smith 1996). Minor codon preference may reflect constraints on RNA structure via ribosomal binding (Eyre-Walker and Bulmer 1993), or perhaps DNA sequence structure (see below for more details of non-translational selection on codon usage).

Mammalian silent site evolution

The best data on silent site selection comes from unicellular organisms such as *E. coli* and small multicellular organisms such as *D. melanogaster* (see above). But there are numerous reasons to study mammals. First, we are mammals ourselves, and as Alexander Pope pointed out "The proper study of mankind is man." Second, there is an enormous, and ever increasing, quantity of mammalian DNA sequence data. Third, the divergence times for mammals are not so great that silent sites reach saturation. Fourth, and of most pertinence to the aims of this thesis, mammals are a good place to look for the limits to selection because large mammals are likely to have small population sizes. From this consideration of population size silent site selection is likely to be weaker in mammals than in enteric bacteria or yeast. Finally, the issue of silent site neutrality is important because the measurement of mutation rates by sequence comparison rests on the assumption of neutrality (Kimura 1983). If mammalian silent sites are not neutral then the use of silent site divergence to estimate mutation rates (as in McVean and Hurst 1997a) is compromised. The measurement of mutation rates is of crucial importance to several problems in evolutionary theory: tests of selection versus mutation, adaptive mutation rate variation across the genome, sex biases in the mutation rate, and the evolution of sex.

Genes across the mammalian genome show great variation in K_S (a 19-fold range over 363 genes was found by Wolfe and Sharp 1993). The comparison of orthologous genes indicates that this variation is gene specific or region specific (Mouchiroud *et al.* 1995). It has been suggested that K_S variation is due to regional mutation biases (Wolfe *et al.* 1989) but later studies showed that the predicted correlation between GC content and K_S does not exist (Bernardi *et al.* 1997; Wolfe and Sharp 1993). However, the use of a "better" maximum likelihood method for substitution rate estimation may have resurrected the finding of a correlation between composition and synonymous substitution rates (see Research Paper 8).

If mutation is not responsible for K_S variation, then selection must be acting on silent sites. Further evidence in favour of weak selection acting on silent sites comes from the overdispersion of the synonymous clock in mammals (Ohta 1995) The fact that the generation time effect is greater for synonymous substitutions implies that nonsynonymous substitutions are subject to stronger selection, and hence that selection on silent sites is weak (Ohta 1993). That the K_A-K_S correlation is stronger than might be expected on the basis of strict neutrality (Ohta and Ina 1995) also suggests selection on silent sites (but see Research Paper 8).

The finding that synonymous substitution rates are similar to rates of divergence in pseudogenes (Wolfe *et al.* 1989) and the non-coding majority of the genome (Wolfe and Sharp 1993) implies that selection on silent sites in mammals is no stronger than selection on non-coding DNA (McVean and Hurst 1997a) (but see Research Papers 9 and 11 for a potential problem in measuring divergence at non-coding sites). Eyre-Walker (1991) has argued that codon usage in mammals is better explained by local mutational biases rather than selection. Circumstantial evidence against silent site selection in mammals comes from the estimate that the selection coefficient against nonoptimal codons multiplied by effective population size is about 0.5 in enteric bacteria (Hartl *et al.* 1994). If the effective population size of mammals is, say, 1000 times smaller than that of *E. coli*, then silent site selection in mammals will be effectively neutral.

Mammalian genes often show a bias towards codons ending in G or C, but it is hard to know whether such codon bias is the result of selection or mutational bias (Li 1997). Selection is implied by stronger codon bias than expected on the basis of background composition (Aota and Ikemura 1986). Even if codon bias was no different from that expected on the basis of compositional variation, that would not rule out selection on silent sites since the compositional patterns in mammals may themselves be due to selection (Bernardi 1995). The finding that compositional patterns differ between sites at different levels of conservation implies that selection does act on DNA base composition (Zoubak *et al.* 1995).

Karlin and Mrazek (1996) have suggested that there are five major processes affecting codon choice in humans. These include selection on translational processes, as well as selection for DNA structure, selection for RNA transcription and stability, mutational biases, and amino acid requirements.

Selection for DNA structure may reflect a need for certain stacking energies, curvatures or superhelicities for stability, or perhaps DNA structure may affect replication and repair processes (Karlin and Mrazek 1996). DNA dinucleotide periodicities of 10-11 bp have been observed in eukaryotic genomes, and have been used to infer selection on DNA structure, since the periodicity is close to the average helical repeat length for DNA in a nucleosome (Widom 1996). However, it appears that the 10-11 bp periodicities are at least partially due to correlations in protein structure (Herzel *et al.* 1998). Karlin and Mrazek (1996) concluded that codon choice in humans is mostly the product of selection on amino acids (for example, GA dinucleotides code for acidic amino acids) and selection for DNA replication and repair. The influence of amino acid constraints and protein structure in mammalian codon usage has also been suggested by Adzhubei *et al.* (1996).

RNA molecules can form complex secondary and tertiary structures strongly related to biological function, and this higher order structure is directed by lower level nucleotide sequence (Conn and Draper 1998; Klaff *et al.* 1996). There is strong evidence for selection on RNA structure in mammalian RNA-coding genes, with structural requirements leading to substitutional correlations between sites which are paired in the RNA secondary structures (P.Higgs, pers. comm.), and rates of nucleotide substitution lower in stems than in loops (Springer *et al.* 1995).

Selection on RNA structure might be thought to proceed by selection for complementary oligonucleotides in close proximity to ensure greater stability at high temperatures, but in prokaryotes such selection is only found for RNA-coding genes and not for genomic base composition (Galtier and Lobry 1997). However, it is unclear whether selection on RNA structure applies to the majority of genes which code for proteins, since mRNA stability in eubacteria and eukaryotes is regulated by RNA structures at the ends of the molecule (Carpousis *et al.* 1999).

The general aim is to understand whether mammalian silent sites in general are affected by selection. Thus cases of selection on silent sites in specific genes (as in Debry and Marzluff 1994) do not resolve the issue and neither would the demonstration that a few silent sites per gene are subject to functional constraints (as suggested by Britten 1993).

Here I give an overview of my findings pertinent to evolution at mammalian silent sites. In Chapter 8 which deals with evolution within introns I describe work on comparing synonymous and intronic substitution rates, the results of which also bear on silent site neutrality.

Repeatable molecular evolution within Igf2r

There is much variation in synonymous substitution rates (K_S) both within (Alvarez-Valin *et al.* 1998) and between (Wolfe and Sharp 1993) mammalian genes. This variation could be the result of stochastic processes (since mutations are rare chance events), or the result of deterministic mutation or selection processes (hence the patterns of synonymous substitution rates make sense). The analysis of the repeatability of patterns of molecular evolution in independent comparisons allows stochastic and deterministic processes to be distinguished (Mouchiroud *et al.* 1995). I investigated the repeatability of molecular evolution for a single large gene *Igf2r*, using the independent pairwise species comparisons of mouse-rat and human-cow (see Research Paper 4 for further details).

Analysis of nonoverlapping windows enabled K_S variation across the gene to be measured for both pairwise comparisons. Using nonparametric statistics the two K_S patterns were significantly similar, and thus K_S across *Igf2r* is repeatable. I believe this result to be the first demonstration of K_S repeatability at such a fine scale (although KS repeatability is suggested by the findings of Comeron and Aguade 1996). The repeatability is not due to methodological artefacts or to the fact that *Igf2r* appears to be a multimeric gene and so I concluded that K_S variation across *Igf2r* is the result of deterministic forces, either mutation or selection.

Two selective explanations of K_S repeatability were considered: mutation rate selection and common constraints. The mutation rate selection hypothesis supposed that silent sites were neutral but that mutation rates could be adaptively reduced in regions of selective importance. The common constraints argument suggested that silent sites might be under translational selection for accuracy and so synonymous divergence would covary with nonsynonymous divergence which is related to selective constraints. The absence of a significant correlation between codon usage bias and K_S went against the common constraints argument, while the absence of a significant

correlations between K_S and either K_A or functional importance was inconsistent with both selection based hypotheses.

The mutational explanation for K_S repeatability suggested that compositional effects might in some way affect mutation rates: the high sequence similarity between the *Igf2r* sequences in the four species would then generate K_S repeatability. Although K_S did not correlate with the crude measure of GC composition, methylation induced CpG => TpG mutations did appear to explain some of the K_S variation. Thus the variation in K_S appeared due to variation in mutable site (CpG) density and the mutability (methylation density) at such sites. The effect of methylation on the molecular evolution of *Igf2r* may be related to the fact that *Igf2r* is genomically imprinted (Neumann *et al.* 1997).

Under the assumption that mammalian silent sites are neutral, as suggested by the tests performed in this study, the repeatability of K_S across *Igf2r* demonstrates mutation rate variation at the intragenic scale. This result changes our view of substitution mutation rate variation, which had previously been demonstrated only at the regional (Casane *et al.* 1997) and chromosomal (McVean and Hurst 1997a) scales.

Tests of codon usage selection on rodent genes

According to the hypothesis of Miyata *et al.* (1987) differences in the mutation rates of genes on different chromosomes are solely due to the proportion of time spent in the male rather than female germline. Hence the comparison of the mutation rates of different chromosomes permits an estimation of the sex bias in the mutation process, thought to be due to the greater number of replications per generation in the male germline (but see Hurst and Ellergren 1998; Li 1997). However synonymous rate comparisons of X chromosomes and autosomes generate different estimates of sex bias from comparisons of X and Y chromosomes. In Chapter 5 and Research Paper 6 I discussed the possibility that this discrepancy is due to different adaptive mutation rates on different chromosomes. Here I review my investigation of the hypothesis that the discrepancy is due to non-neutral evolution, that X-linked genes have reduced K_S due to the greater efficiency of hemizygous selection against recessive deleterious mutations (see Research Paper 6).

I measured substitution rates and codon bias for a set of 297 autosomal and 37 X-linked pairs of confirmed mouse-rat orthologs, and then examined the evidence for codon usage selection in a number of ways, using two different classes of codon usage bias measures. The Effective Number of Codons (ENC) (Wright 1990) can vary from 20 with one codon for each amino acid to 61 with all synonymous codons used equally, and hence negatively correlates with codon bias. In order to account for compositional effects I adapted the ENC method to generate an alternative measure called CUBRE, Codon Usage Bias Relative to Expected. Neither the ENC nor the CUBRE data supported either of two predictions of codon usage selection on silent sites: (1) synonymous substitution rates should correlate negatively with codon bias, and (2) since K_S is lower on the X chromosome than on autosomes the codon bias should be greater on the X

chromosome. However, the CUBRE data did show more codon usage bias than expected on the basis of compositional bias.

Given that mutational biases might have generated such an excess of codon bias, I used Eyre-Walker's (1991) X^2 tests of heterogeneity which account for the potential mutational biases of neighbouring nucleotides (Bulmer 1986). Codon usage heterogeneity was weak with only one of the four groups of codons (the AGA group) showing significantly more biased genes than expected by chance. Those genes which showed significant heterogeneity in codon usage did not have lower synonymous substitution rates, as expected on the basis of selection on codon usage. The heterogeneity of the AGA group was further examined by measuring the heterogeneity in different reading frames, with the results favouring long range mutational biases (Bulmer 1990) rather than selection. Codon usage heterogeneity did not show the negative correlation with synonymous substitution rates expected on the basis of selection for codon usage.

An alternative test examined the basis of the argument that K_S is lower on the X chromosome because of selection on silent sites and hemizygous expression. Such an explanation assumes that the majority of mutations are deleterious since advantageous recessive mutations would probably spread faster on the X chromosome (Charlesworth *et al.* 1987). If most mutations are deleterious, and if the fitness distribution of mutations is the same for synonymous and nonsynonymous mutations, then the X chromosome should have an even lower K_A than expected on the basis of its low K_S. Such a pattern was not found and so this test suggested that not all mutations are deleterious, which destroys the basis for suggesting that the low K_S on the X chromosome is the result of selection on synonymous mutations.

In this study I found no clear evidence of selection on synonymous sites in rodents. Codon bias did not show the predicted correlation with K_S, and the comparison of the K_S and K_A data indicates that the low K_S on the X chromosome actually constitutes an argument against selection on silent sites. The only evidence in favour of codon usage selection was the finding that some measures of codon usage bias were higher than expected, but this could well have been due to mutational biases. That most of the excessive bias disappears when neighbouring bases are controlled for is consistent with this belief.

Substitution patterns indicate selection on silent sites

Selection on silent sites provides a potential explanation for the finding that the K_A-K_S correlation in mammals is stronger than expected on the basis of the neutral theory and silent site neutrality (as shown by Ohta and Ina 1995). I have investigated the K_A-K_S correlation in rodents and have found evidence that the strength of the correlation may be partly due to selection on silent sites which causes an excess of tandem substitutions (see Research Paper 8).

The strength of the K_A-K_S correlation is consistent with silent site neutrality and the neutral theory if one uses maximum likelihood methods to estimate substitution rates (but not if uses

algorithmic methods). However, tandem substitutions account for much of the strength of the correlation irrespective of rate estimation method, and so one would like to explain why there are so many tandem substitutions.

There is an excess of tandem substitutions over the null expectation based on independent substitutions. This excess of tandem substitutions can be explained in one of two ways: either doublet mutations or correlated selection pressures affecting adjacent bases. I investigated these alternatives by the analysis of substitution patterns (see Research Paper 8 for details of the method). These patterns indicated that doublet mutations could not be responsible and so, by elimination of the alternatives, selection on silent sites must be responsible.

Given that I had some evidence of selection on silent sites in rodents I attempted to discover the nature of that selection. Employing a test based on substitution patterns, I found no evidence for selection on codon usage. The use of maximum likelihood (but not algorithmic) methods revealed correlations between K_S and base composition parameters, consistent with but certainly not strong evidence for selection affecting base composition. I also found K_S to be reduced at the start of genes, a result which has also been found in *E. coli* and interpreted in that case as evidence for selection on RNA structure. This study represents the first evidence of selection acting on silent sites in many mammalian genes.

Research Paper 8. The effect of tandem substitutions on the correlation between synonymous and nonsynonymous rates in rodents

Nick Smith And Laurence Hurst (1999)
Genetics **153** 1395-1402.

The Effect of Tandem Substitutions on the Correlation Between Synonymous and Nonsynonymous Rates in Rodents

Nick G. C. Smith and Laurence D. Hurst

Department of Biology and Biochemistry, University of Bath, Bath B42 7AY, United Kingdom

Manuscript received March 31, 1999
Accepted for publication August 2, 1999

ABSTRACT

Nonsynonymous substitutions in DNA cause amino acid substitutions while synonymous substitutions in DNA leave amino acids unchanged. The cause of the correlation between the substitution rates at nonsynonymous (K_A) and synonymous (K_S) sites in mammals is a contentious issue, and one that impacts on many aspects of molecular evolution. Here we use a large set of orthologous mammalian genes to investigate the causes of the K_A-K_S correlation in rodents. The strength of the K_A-K_S correlation exceeds the neutral theory expectation when substitution rates are estimated using algorithmic methods, but not when substitution rates are estimated by maximum likelihood. Irrespective of this methodological uncertainty the strength of the K_A-K_S correlation appears mostly due to tandem substitutions, an excess of which is generated by substitutional nonindependence. Doublet mutations cannot explain the excess of tandem synonymous-nonsynonymous substitutions, and substitution patterns indicate that selection on silent sites is the likely cause. We find no evidence for selection on codon usage. The nature of the relationship between synonymous divergence and base composition is unclear because we find a significant correlation if we use maximum-likelihood methods but not if we use algorithmic methods. Finally, we find that K_S is reduced at the start of genes, which suggests that selection for RNA structure may affect silent sites in mammalian protein-coding genes.

THE nature of the relationship between nonsynonymous and synonymous substitution rates pertains to many aspects of molecular evolution in mammals (INA 1996b). A link between the processes of evolution at synonymous and nonsynonymous sites may be due to selection on synonymous sites (see below). Selection on silent sites would affect the selectionist-neutralist debate, for example, providing a potential explanation for the overdispersion of synonymous substitution rates (as shown by OHTA 1995), and would call into question the practice of using silent site comparison to study the evolution of mutation rates (as in MCVEAN and HURST 1997a).

Several studies have reported a highly significant positive correlation between the synonymous substitution rates (K_S) and the nonsynonymous substitution rates (K_A) of mammalian genes (WOLFE and SHARP 1993; MOUCHIROUD *et al.* 1995; MAKALOWSKI and BOGUSKI 1998b). The K_A-K_S correlation also appears to hold within some mammalian genes (ALVAREZ-VALIN *et al.* 1998).

In this article we investigate a variety of explanations for the intergenic K_A-K_S correlation in mammals, specifically in the comparison between mouse and rat. A number of hypotheses for the K_A-K_S correlation exist (for example see LI 1997). An attractive null hypothesis for the K_A-K_S correlation is the neutral theory explanation, which supposes that genes differ in mutation rates, that all synonymous changes are neutral, and that a variable proportion of nonsynonymous changes are neutral (OHTA and INA 1995). This null hypothesis can explain the existence of a K_A-K_S correlation in mammals but appears unable to explain why the correlation is so strong (OHTA and INA 1995). We confirm this result using an improved data set and algorithmic rate estimation methods, but we also show that the K_A-K_S correlation is consistent with the neutral prediction if one uses maximum likelihood (ML) to estimate substitution rates (see RESULTS). Thus methodological bias may have led to previous overestimates of the strength of the K_A-K_S correlation.

Despite the fact that the strength of the K_A-K_S correlation may be consistent with silent site neutrality, patterns of substitutions indicate that selection may well be acting on silent sites. In particular, the strength of the K_A-K_S correlation appears in large part to be due to an excess of tandem substitutions caused by substitutional nonindependence.

Synergy between synonymous and nonsynonymous substitutions, such that one type of substitution increases the likelihood of the other, would increase the K_A-K_S correlation. Such substitutional nonindependence could be the result of either selection or muta-

Corresponding author: Nick G. C. Smith, School of Biological Sciences, University of Sussex, Brighton BN1 9QG, United Kingdom.
E-mail: n.g.c.smith@sussex.ac.uk

tion. Purifying selection might act on both nonsynonymous and synonymous sites (Ina 1996a), or nonsynonymous substitutions might cause positive selection on subsequent synonymous substitutions (Lipman and Wilbur 1985). Alternatively, a single mutational event might affect both synonymous and nonsynonymous sites simultaneously as with doublet mutations (Wolfe and Sharp 1993). (Note on terminology: we use "doublet" to refer to a supposed mutational event affecting adjacent bases and "tandem" to apply to observed adjacent substitutions.)

It is also possible to envisage a hybrid selection-mutation model in which a correlation between the mutation rate and nonsynonymous constraints causes an increase in the K_A-K_S correlation (Ina 1996a). Such a hybrid explanation is supported by theoretical (Kondrashov 1995) and empirical (McVean and Hurst 1997b; Smith and Hurst 1999) studies of the evolution of mutation rates.

MATERIALS AND METHODS

Selection of protein coding sequences: A list of 470 genes in mouse, rat, and human, with orthology confirmed using HOVERGEN 19 (Duret *et al.* 1994), was obtained from Makalowski and Boguski (1998a). Only genes with complete protein-coding sequence available in a single GenBank/EMBL record were used, leaving 432 three-species comparisons.

Preparation of alignments: Alignments were performed using the GCG (1994) and EGCG (Rice 1997) packages at HGMP (http://www.hgmp.mrc.ac.uk/). FETCH was used to extract sequences from databases, and GENETRANS was used to extract and combine exons automatically. Protein alignments were performed using CLUSTALW (Thompson *et al.* 1994). Then the DNA alignments were recreated from the protein alignments and the original DNA sequences using the program MRTRANS (written by W. Pearson and available at HGMP).

ML analysis: The ML package PAML (Yang 1997) was used to reconstruct ancestral sequences and to estimate substitution rates. We used the program BASEML to reconstruct ancestral sites, with the gene tree defined as [(mouse, rat), human], with no rate variation between sites, and with the REV model of evolution. Ancestral sequence reconstruction was carried out by ML, rather then parsimony, for two reasons: ML allows the reconstruction of all sites, and parsimony is biased when base composition is skewed (Eyre-Walker 1998).

The program CODEML, under a codon-based model of evolution (Goldman and Yang 1994), was used to estimate K_A and K_S. Using PAML version 2.0 the following parameter settings were used: seqtype = 1, codon-based model; runmode = −2, estimate K_A and K_S rates; CodonFreq = 3, codon frequencies used as free parameters; additionally, no rate variation was allowed.

Algorithmic rate estimation: Substitution rates were also estimated from sequence alignments using algorithmic methods developed by Moriyama and Powell (1997). Tamura's (1992) multiple hits correction method was used in conjunction with Li's (1993) method to calculate K_A and K_S. The substitution rates at fourfold synonymous sites, K_4, were also estimated using the algorithmic method of Tamura and Nei (1993). Estimates of K_4 are expected to be more reliable than estimates of K_S, which have to combine the rates of sites of different degeneracies.

With regard to the differences between the algorithmic and PAML rate estimation methods, the algorithmic methods gave similar results to PAML using CodonFreq = 1, codon frequencies calculated from average nucleotide frequencies. But with PAML using CodonFreq = 2, codon frequencies calculated from average nucleotide frequencies at the three codon positions, and PAML using CodonFreq = 3, codon frequencies as free parameters, the PAML and algorithmic estimates differed with regard to the strengths of the K_A-K_S and the K_S-composition correlations (data not shown, but see RESULTS for a comparison of the algorithmic estimates and PAML estimates using CodonFreq = 3).

Measurement of substitutional nonindependence: To analyze lineage-specific substitution patterns, we used mouse, rat, and human orthologs to reconstruct ancestral sequences (see above) and compared present-day sequences to their most recent ancestral node. The mouse and rat lineage-specific substitution patterns were combined.

The measurement of substitution patterns proceeded as follows. Substitutions between two sequences were designated as either fully synonymous (syn) or fully nonsynonymous (nonsyn) or mixed (part syn and part nonsyn), following the method of Li *et al.* (1985). All substitutions within 100 bp of every other substitution were investigated, and the totals of all substitution pairs a certain distance apart were noted (if one or both of the substitutions was mixed the necessary weightings were applied, and indels were ignored). Three classes of substitution pairs were investigated: syn-syn, syn-nonsyn, and nonsyn-nonsyn.

Simulated substitution sequences were generated under the assumption of independent substitutions. Simulated sequences were the same length as the real sequences and were generated according to the codon position-specific synonymous and nonsynonymous substitution rates of the real sequences so that the substitution rates of the simulations were the same as those of the real sequences. The same substitution pattern analysis was performed on the simulated sequences as on the real sequences. For each sequence considered the substitution patterns of the real sequence were compared against those of 500 simulated sequences.

Statistics describing the difference between the real and simulated substitution patterns were calculated for all three substitution pair classes. The greater the difference between the real and simulated substitution patterns the greater the nonindependence between real substitutions, and thus we term our statistic substitutional nonindependence (SNI). The numbers of real cases (r) were summed for all N sequences, and for each simulation run the numbers of simulated occurrences (s) were summed for all sequences. SNI is given by the number of simulation runs for which the real total was greater than the simulated total, so for 500 simulations per sequence we have the formula

$$\mathrm{SNI} = \sum_{j=1}^{500}\left(\sum_{i=1}^{N} r_i > \sum_{i=1}^{N} s_{ij}\right).$$

Under the null assumption of no difference between real and simulated substitution patterns, the expected value of SNI is 250. Using the normal distribution as an approximation to the binomial, we find the one-tailed 95% upper confidence limit to be 268. If we apply the Bonferroni correction for considering 100 different substitution pair distances (as described on page 240 in Sokal and Rohlf 1995), the upper confidence limit is 286.

As an aid to visualization of substitutional nonindependence, we have also provided plots of substitution pair class separation against the statistic real over simulated (ROS),

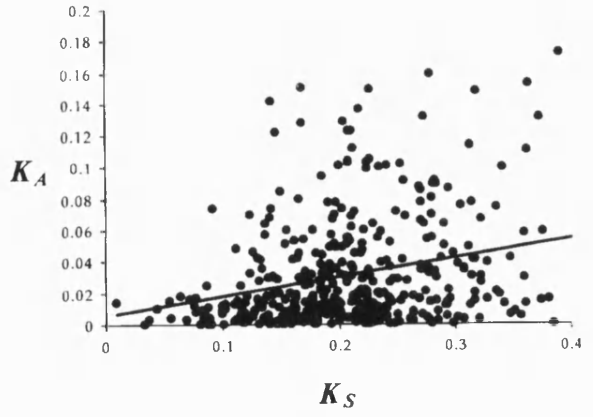


FIGURE 1.—K_A plotted against K_S for 432 mouse-rat genes. Substitution rates were estimated using maximum-likelihood methods. The linear regression line is shown.

which increases from unity upward as substitutional nonindependence increases, and which is defined as

$$\mathrm{ROS} = \frac{(1/N)\Sigma_{i=1}^{N} r_i}{(1/500N)\Sigma_{j=1}^{500}\Sigma_{i=1}^{N} s_{ij}}.$$

RESULTS

The K_A-K_S correlation is consistent with neutral theory: Using ML rate estimation and a large set of orthologous mouse-rat genes (see MATERIALS AND METHODS and Figure 1), we estimated the K_A-K_S correlation coefficient by rank correlation followed by the z-transformation (SOKAL and ROHLF 1995). Contrary to the suggestion that a significant K_A-K_S correlation results from the inclusion of paralogs (HUGHES and YEAGER 1997), and in agreement with previous findings (MOUCHIROUD *et al.* 1995; MAKALOWSKI and BOGUSKI 1998b), we find a highly significant positive correlation between K_A and K_S ($P < 0.0001$) using both algorithmic and ML rate estimation methods.

In addition we calculated the K_A-K_S correlation coefficients predicted by the neutral theory explanation as given by OHTA and INA (1995). In agreement with their results, we find that the neutral theory is unable to explain the strength of the observed K_A-K_S correlation if algorithmic rate estimation methods are used, with the observed correlation coefficient R greater than the expected correlation coefficient ρ ($R = 0.411$ and $\rho = 0.270$). Statistical testing is difficult because the variance of ρ is not theoretically tractable (OHTA and INA 1995), but simulations have shown that findings of $R \gg \rho$ can be explained by pure chance (INA 1996a).

However, in contrast to the results of OHTA and INA (1995), we find that the neutral theory is consistent with the strength of the observed K_A-K_S correlation if ML rate estimation methods are used, with R less than but similar to ρ ($R = 0.275$ and $\rho = 0.343$). The evolutionary model specified in PAML is more general than that of the algorithmic method, which might lead one to conclude that the PAML rate estimates are probably more reliable. However, the PAML rate estimates should not be considered perfect: standard errors, required to predict ρ, are estimated using the normal approximation to the likelihood curve; and the model of evolution makes no allowance for rate variation between sites. There is also the question of whether pairwise sequence comparisons provide enough data for the ML approach to provide unbiased estimates. We conclude that it is unclear which of the algorithmic or ML approaches is more reliable and thus can only note the methodological sensitivity of the strength of the K_A-K_S correlation relative to the neutral theory prediction.

The importance of tandem substitutions: The influence of tandem substitutions was investigated using ML rate estimation (similar results were obtained using algorithmic methods). If tandem substitutions were ignored, the expected correlation coefficient considerably exceeded the observed correlation coefficient ($R = 0.046$ and $\rho = 0.349$); thus tandem substitutions appear to make a large contribution to the strength of the K_A-K_S correlation. Upon removal of tandem substitutions the ratio of the expected correlation coefficient to the observed correlation coefficient changes from 1.25 to 7.59, a sixfold increase.

If only those genes with no tandem substitutions were considered ($N = 67$), the K_A-K_S correlation was zero, considerably below the neutral expectation ($R = 0$ and $\rho = 0.344$). This result suggests that the K_A-K_S correlation is generated almost exclusively by tandem substitutions, although this interpretation should be treated with caution as the genes with no tandem substitutions were atypically short and slowly evolving (data not shown).

Substitutional nonindependence mainly affects adjacent bases: The K_A-K_S correlation is strengthened if there is substitutional nonindependence between synonymous and nonsynonymous sites (see Introduction). The effect of tandem substitutions on the K_A-K_S correlation implies nonindependence between adjacent substitutions; but does substitutional nonindependence occur at other distances? We measured the nonindependence between syn-nonsyn pairs of substitutions at all pair separation distances from 1 to 100 bases (see MATERIALS AND METHODS). If all substitutions are considered, then substitutional nonindependence appears to operate at a variety of distances: 80 of the 100 syn-nonsyn pairs have highly significant SNI values ($P < 0.05$ with Bonferroni correction). The ROS plot (Figure 2) shows high levels of substitutional nonindependence for the syn-nonsyn pairs, with ROS values tending to decrease as the distance between the two substitutions increases (note that tandem syn-nonsyn substitutions give the highest ROS value).

To check whether substitutional nonindependence really exists beyond effects between adjacent bases, we investigated the effect of the removal of tandem substitu-

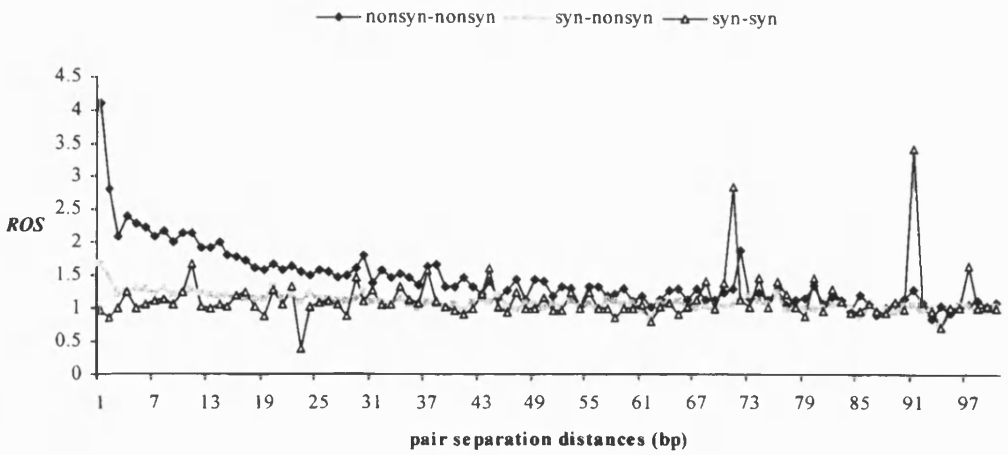


FIGURE 2.—The ROS plots of substitutional nonindependence for the rodent lineages. Values are given for all three classes of substitution and for all pair separation distances from 1 to 100 bp (see MATERIALS AND METHODS).

tions on substitution patterns. The resultant change in patterns of substitutional nonindependence is striking (compare Figures 2 and 3). Not a single syn-nonsyn pair yielded a significantly high SNI value ($P > 0.05$ without Bonferroni correction). These results imply that whatever process (selection or mutation) is responsible for the nonindependence of syn-nonsyn substitutions, then that process is mainly acting on adjacent bases and causing an excess of tandem substitutions.

The excess of tandem substitutions is not due to doublet mutations: If mutational processes are sufficient to explain the excess of tandem substitutions without recourse to selection, then synonymous changes are neutral and doublet mutations are responsible for the excess of tandem substitutions. From these assumptions we can predict an excess of neighboring syn-syn pairs. But the SNI value for neighboring syn-syn pairs is 143, which is lower than the null expectation of 250. This means that either doublet mutations do not occur or that synonymous doublet mutations are subject to purifying selection. Either way, we can conclude that mutation alone is unable to explain the excess of tandem substitutions. Hence by elimination we are left with a selective explanation for the excess of tandem substitutions.

Selection on silent sites is demonstrated by patterns of substitutional nonindependence: We have shown that synonymous-nonsynonymous substitutional nonindependence does not appear to exist beyond the interactions of adjacent bases. Given that we have also provided evidence against doublet mutations, we have no reason to believe in any form of mutational nonindependence. If we make the assumption that mutation does not differentiate between synonymous and nonsynonymous sites, then we can conclude that any differences in substitutional nonindependence between the three classes of substitution pairs (syn-syn, syn-nonsyn, and nonsyn-nonsyn) must be due to selection.

The different types of substitution pairs do indeed show significantly different levels of substitutional nonindependence. For each class of pairs, 100 different measures of SNI were obtained, corresponding to all the pair separation distances from 1 to 100 bp. Out of a possible maximum of 100, 96 of the nonsyn-nonsyn pair classes, 80 of the syn-nonsyn pair classes, and 69 of the syn-syn pair classes have highly significant SNI values ($P < 0.05$ with Bonferroni correction). The ROS plots (Figure 2) show the same pattern of substitutional nonindependence decreasing in the order of nonsyn-nonsyn, syn-nonsyn, and syn-syn. Both the nonsyn-nonsyn

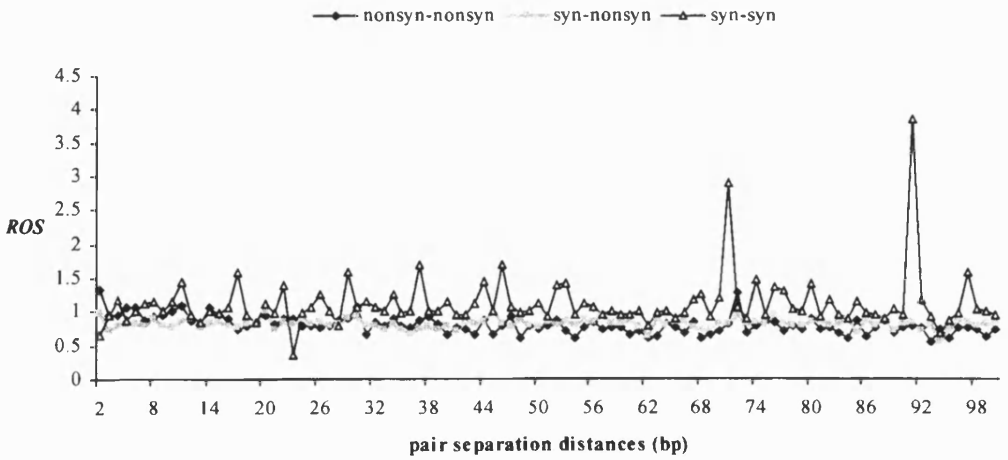


FIGURE 3.—The ROS plots of substitutional nonindependence for the rodent lineages after removal of tandem substitutions. Compare with Figure 2.

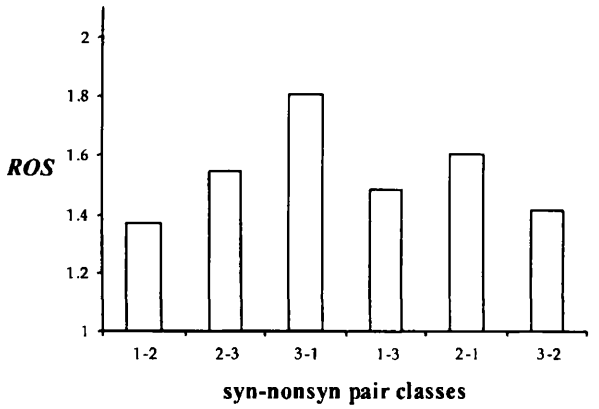


FIGURE 4.—The ROS measures of substitutional nonindependence for six classes of syn-nonsyn pairs as defined by the codon positions of the substitutions. If selection acts on codon usage, then those pairs contained within a single codon (1-2, 2-3, and 1-3) should have higher ROS values than the other pairs (3-1, 2-1, and 3-2).

(Mann-Whitney *U*-test, $P < 0.0001$) and syn-nonsyn (Mann-Whitney *U*-test, $P = 0.04$) pair classes show significantly greater substitutional nonindependence than the syn-syn pair class.

These results do not appear to be the result of unreliable ancestral sequence reconstruction, because qualitatively identical results are obtained from the mouse-rat interspecies comparison as from the lineage-specific comparisons (data not shown). Therefore selection appears to be operating on silent sites, though we accept that our conclusion is based on an assumption concerning the nature of the mutational process. We now attempt to discern the precise nature of the selection on silent sites.

Selection for major codon usage: If selection acts to favor major codon usage (AKASHI and EYRE-WALKER 1998), then substitutional nonindependence should be greater for pairs of substitutions within codons than pairs of substitutions between codons.

The syn-nonsyn substitution pairs at distances of 1 and 2 bp were both divided into three classes according to the codon positions of the substitutions. The pairs 1 bp apart were classified as 1-2, 2-3, and 3-1. Both 1-2 and 2-3 represent a pair of substitutions within a codon, while 3-1 invokes substitutions in adjacent codons. Similarly, the pairs 2 bp apart were classified as 1-3, 2-1, and 3-2. In this case only 1-3 comprises substitutions within a codon, while both 2-1 and 3-2 involve substitutions in adjacent codons.

All six substitution pair classes show highly significant SNI values ($P < 0.05$ with Bonferroni correction), and thus the SNI data are equivocal on the issue of selection for codon usage. The ROS data are contrary to predictions based on selection for codon usage: ROS is greater in the 3-1 class than in the 1-2 and 2-3 classes, and ROS in the 1-3 class is intermediate between that in the 2-1 and 3-2 classes (see Figure 4). Our finding of no evidence in favor of selection for major codon usage in mammals supports previous studies (EYRE-WALKER 1991; SMITH and HURST 1999).

Selection for base composition: The relationships between synonymous substitution rates and a number of compositional characters were examined to test predictions of specific selective pressures. Significant correlations would be consistent with selection acting directly on base composition or a link between selection and other characters that correlate with composition (such as recombination; EYRE-WALKER 1993). However, this test is not capable of providing strong evidence in favor of selection, because K_S-composition correlations could be the result of mutation rather than selection.

As with the K_A-K_S correlation, the alternative methods of rate estimation yield different results. With the algorithmic method K_S does not correlate strongly with either *GC4* (G plus C content at fourfold degenerate sites; $R = 0.008$), *A4* ($R = -0.03$), *C4* ($R = -0.025$), *G4* ($R = 0.071$), or *T4* ($R = -0.007$). Using the more reliable algorithmic measure of K_4 we also find no correlation between synonymous divergence and base composition (*GC4* and K_4; $R = 0.002$; see Figure 5). However, with PAML we find significant correlations ($P < 0.0001$) for all compositional parameters: *GC4* ($R = 0.258$; see Figure 5), *A4* ($R = -0.264$), *C4* ($R = 0.187$), *G4* ($R = 0.247$), and *T4* ($R = -0.206$).

These differences between the methods are all the more surprising when one considers that, as one would expect, the alternative measures of synonymous divergence are highly significantly correlated ($R \sim 0.9$). Given that we are unable to choose between algorithmic and ML methods (see above), these data are equivocal on the issue of selection on silent sites (for evidence of selection on the base composition of mammalian silent sites, see EYRE-WALKER 1999). However, our results are pertinent to the debate as to whether there is a relationship between K_S and base composition. The existence of a significant correlation was originally suggested by WOLFE *et al.* (1989) on the basis of a fairly small sample. BERNARDI *et al.* (1997) subsequently showed that the inverted *V* distribution obtained by WOLFE *et al.* (1989) was at least partially due to rate estimate biases (see PESOLE *et al.* 1995). However, our ML results suggest a linear relationship between *GC4* and K_S (see Figure 5), which cannot be so easily explained by methodological biases.

Selection for RNA structure: Selection on RNA structure has been proposed as an explanation for the reduced K_S at the start of protein-coding enterobacterial genes, with an open structure thought to favor ribosome binding (EYRE-WALKER and BULMER 1993). We have found a similar pattern in our set of mammalian genes (see Figure 6). For all 354 genes with mouse-rat alignments longer than 600 bp, K_S was estimated using algorithmic methods for five regions of the gene: the whole gene and the first four nonoverlapping sections of 50

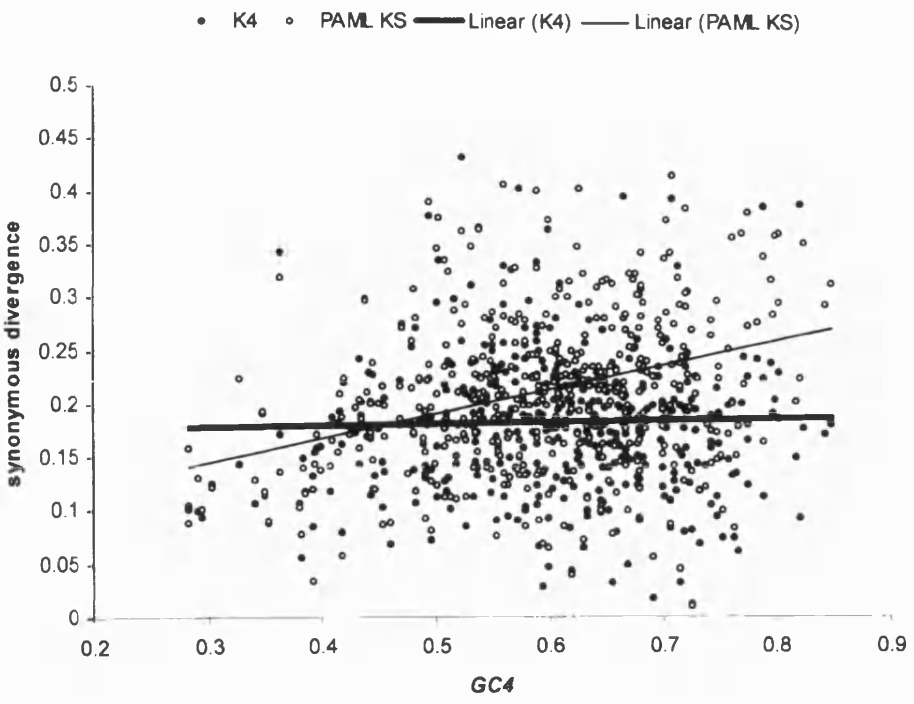


FIGURE 5.—Base composition at fourfold degenerate sites (*GC4*) plotted against synonymous divergence for 432 mouse-rat genes. Two measures of synonymous divergence are shown: PAML K_S is a maximum-likelihood estimate while K_4 is an algorithmic estimate (see MATERIALS AND METHODS). The linear regression lines show a significant relationship between PAML K_S and *GC4* but not between K_4 and *GC4*.

codons. The first 50 codons at the start of the gene have a significantly low K_S in comparison to both the whole gene (Mann-Whitney *U*-test, $P < 0.0001$) and three subsequent 50-codon blocks (Mann-Whitney *U*-tests, $P = 0.0019$, $P = 0.0081$, $P = 0.0047$). These findings provide us with suggestive, although by no means conclusive, evidence that silent sites in mammals are affected by selection.

It is thought that longer mRNAs have a lower density of longer stem loops, and so selection on RNA structure is predicted to decrease with increasing gene length (COMERON and AGUADE 1996). We find no correlation between gene length and K_S for either rate estimation method, though we note that this appears to be a weak test of selection.

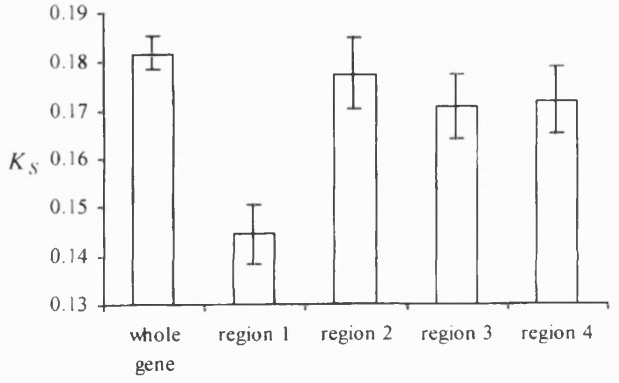


FIGURE 6.—The first 50 codons of a gene (region 1) have a low K_S relative to the whole gene and the three subsequent 50-codon nonoverlapping sections (regions 2 to 4). The error bars indicate the SEs of the means. Rates were estimated using algorithmic methods.

DISCUSSION

With a ML approach to rate estimation, the rodent K_A-K_S correlation coefficient is consistent with the neutral theory, but using an algorithmic approach the correlation is stronger than expected. Despite such methodological uncertainty we have found strong evidence to suggest that the excess of tandem substitutions generated by substitutional nonindependence contributes to the strength of the rodent K_A-K_S correlation coefficient. The removal of tandem substitutions reduces the K_A-K_S correlation coefficient by a factor of six, and there exists no K_A-K_S correlation for those genes that do not contain tandem substitutions. Substitutional nonindependence between adjacent bases, the process that generates the excess of tandem substitutions, appears to be the dominant form of substitutional nonindependence.

What causes the excess of tandem substitutions that contribute to the K_A-K_S correlation? Is it selection or mutation? We demonstrate that the mutational explanation fails due to a lack of evidence for doublet mutations, which means that selection must be responsible for the excess of synonymous-nonsynonymous tandem substitutions. Our analysis of the substitution patterns of the different pair classes also supports the notion of silent site selection, and encourages us to investigate the form of selection acting on silent sites. It might be argued that our finding of substitutional nonindependence caused by selection is inconsistent with our finding using ML methods that the K_A-K_S correlation is consistent with neutrality, but the neutral prediction should remain reasonably accurate as long as the proportion of silent sites affected by selection is low. Although tandem substitutions are contributing greatly to the K_A-K_S correla-

tion, selection may generate a relatively small excess of tandems above those predicted on the basis of neutrality.

By examining substitution patterns we have provided evidence against selection acting on codon usage. We have found that the existence of correlations between K_S and base composition depends on rate estimation methodology and offers no clue as to whether selection via base composition acts on silent sites. There is no correlation between K_S and gene length, but selection on RNA structure is consistent with our finding that K_S is reduced at the start of mammalian genes. Although further work is clearly required to examine this supposition, we suggest that selection on RNA structure is a possible explanation for the strong syn-syn substitutional nonindependence at distances of 71 and 91 bp (see Figure 2).

What are the implications of our results with respect to mammalian molecular evolution? We have found three reasons to believe that silent sites in mammals are subject to selection: (i) mutation cannot explain the excess of syn-nonsyn tandem substitutions, therefore selection is responsible by elimination; (ii) a comparison of the levels of substitutional nonindependence of the syn-syn, syn-nonsyn, and nonsyn-nonsyn classes of substitution pairs appears to indicate the effects of selection; and (iii) low K_S at the start of genes is consistent with selection on RNA structure. Although arguments (ii) and (iii) are by no means certain, we consider reason (i) to provide strong evidence for silent site selection.

Selection on silent sites can explain the overdispersion of silent sites in mammals (as in OHTA 1995). But does silent site selection necessarily invalidate those studies of the evolution of the mutation rate in mammals, which assume that silent sites are neutral and hence that K_S can be used as an unbiased estimator of the mutation rate (as in MCVEAN and HURST 1997a)? Although we have found evidence of selection on silent sites we still believe that K_S provides the best available estimate of the mutation rate. First, K_S values before and after the removal of tandem substitutions are highly significantly correlated (using PAML, $R = 0.927$ and $P < 0.00001$). Second, tests of adaptive mutation rates hold both before and after the removal of tandem substitutions (SMITH and HURST 1999). Third, there is a practical argument in favor of using K_S, which is that the alternative way to estimate mutation rates is to use noncoding DNA sequence data, the alignment of which is problematic (SMITH and HURST 1998).

The authors thank Ziheng Yang, Yasuo Ina, Adam Eyre-Walker, Paul Higgs, and Jonathan Slack. L.D.H. is funded by the Royal Society.

LITERATURE CITED

AKASHI, H., and A. EYRE-WALKER, 1998 Translational selection and molecular evolution. Curr. Opin. Genet. Dev. **8**: 688–693.

ALVAREZ-VALIN, F., K. JABBARI and G. BERNARDI, 1998 Synonymous and nonsynonymous substitutions in mammalian genes: intragenic correlations. J. Mol. Evol. **46**: 37–44.

BERNARDI, G., D. MOUCHIROUD and C. GAUTIER, 1997 Isochores and synonymous substitutions in mammalian genes, pp. 137–168 in *DNA and Protein Sequence Analysis*, edited by M. J. BISHOP and C. J. RAWLINGS. Oxford University Press, Oxford.

COMERON, J. M., and M. AGUADE, 1996 Synonymous substitutions in the Xdh gene of *Drosophila*—heterogeneous distribution along the coding region. Genetics **144**: 1053–1062.

DURET, L., D. MOUCHIROUD and M. GOUY, 1994 Hovergen—a database of homologous vertebrate genes. Nucleic Acids Res. **22**: 2360–2365.

EYRE-WALKER, A., 1991 An analysis of codon usage in mammals: selection or mutation bias? J. Mol. Evol. **33**: 442–449.

EYRE-WALKER, A., 1993 Recombination and mammalian genome evolution. Proc. R. Soc. Lond. Ser. B **252**: 237–243.

EYRE-WALKER, A., 1998 Problems with parsimony in sequences of biased base composition. J. Mol. Evol. **47**: 686–690.

EYRE-WALKER, A., 1999 Evidence of selection on silent site base composition in mammals: potential implications for the evolution of isochores and junk DNA. Genetics **152**: 675–683.

EYRE-WALKER, A., and M. BULMER, 1993 Reduced synonymous substitution rate at the start of enterobacterial genes. Nucleic Acids Res. **21**: 4599–4603.

GCG, 1994 *Program Manual for the Wisconsin Package, Version 8*. Genetics Computer Group, Madison, WI.

GOLDMAN, N., and Z. YANG, 1994 A codon-based model of nucleotide substitution for protein-coding DNA sequences. Mol. Biol. Evol. **11**: 725–736.

HUGHES, A. L., and M. YEAGER, 1997 Comparative evolutionary rates of introns and exons in murine rodents. J. Mol. Evol. **45**: 125–130.

INA, Y., 1996a Correlation between synonymous and nonsynonymous substitutions and variation in synonymous substitution numbers, pp. 105–113 in *Current Topics on Molecular Evolution*, edited by M. NEI and N. TAKAHATA. Institute of Molecular Evolutionary Genetics, Penn State University, University Park, PA and The Graduate University for Advanced Studies, Hayama, Japan.

INA, Y., 1996b Pattern of synonymous and nonsynonymous substitutions: an indicator of mechanisms of molecular evolution. J. Genet. **75**: 91–115.

KONDRASHOV, A. S., 1995 Modifiers of mutation-selection balance: general-approach and the evolution of mutation-rates. Genet. Res. **66**: 53–69.

LI, W. H., 1993 Unbiased estimation of the rates of synonymous and nonsynonymous substitution. J. Mol. Evol. **36**: 96–99.

LI, W. H., 1997 *Molecular Evolution*. Sinauer Associates, Sunderland, MA.

LI, W. H., C. I. WU and C. C. LUO, 1985 A new method for estimating synonymous and nonsynonymous rates of nucleotide substitution considering the relative likelihood of nucleotide and codon changes. Mol. Biol. Evol. **2**: 150–174.

LIPMAN, D. J., and W. J. WILBUR, 1985 Interaction of silent and replacement changes in eukaryotic coding sequences. J. Mol. Evol. **21**: 161–167.

MAKALOWSKI, W., and M. S. BOGUSKI, 1998a Evolutionary parameters of the transcribed mammalian genome: an analysis of 2,820 orthologous rodent and human sequences. Proc. Natl. Acad. Sci. USA **95**: 9407–9412.

MAKALOWSKI, W., and M. S. BOGUSKI, 1998b Synonymous and nonsynonymous substitution distances are correlated in mouse and rat genes. J. Mol. Evol. **47**: 119–121.

MCVEAN, G. T., and L. D. HURST, 1997a Evidence for a selectively favourable reduction in the mutation rate of the X chromosome. Nature **386**: 388–392.

MCVEAN, G. T., and L. D. HURST, 1997b Molecular evolution of imprinted genes: no evidence for antagonistic coevolution. Proc. R. Soc. Lond. Ser. B **264**: 739–746.

MORIYAMA, E. N., and J. R. POWELL, 1997 Synonymous substitution rates in Drosophila: mitochondrial versus nuclear genes. J. Mol. Evol. **45**: 378–391.

MOUCHIROUD, D., C. GAUTIER and G. BERNARDI, 1995 Frequencies of synonymous substitutions in mammals are gene-specific and correlated with frequencies of nonsynonymous substitutions. J. Mol. Evol. **40**: 107–113.

OHTA, T., 1995 Synonymous and nonsynonymous substitutions in mammalian genes and the nearly neutral theory. J. Mol. Evol. **40**: 56–63.

OHTA, T., and Y. INA, 1995 Variation in synonymous substitution rates among mammalian genes and the correlation between synonymous and nonsynonymous divergences. J. Mol. Evol. **41:** 717–720.

PESOLE, G., G. DELLISANTI, G. PREPARATA and C. SACCONE, 1995 The importance of base composition in the correct assessment of genetic distance. J. Mol. Evol. **41:** 1124–1127.

RICE, P., 1997 *Program Manual for the EGCG Package.* The Sanger Centre, Hinxton Hall, Cambridge, CB10 1RQ, England.

SMITH, N. G. C., and L. D. HURST, 1998 Sensitivity of patterns of molecular evolution to alterations in methodology: a critique of Hughes and Yeager. J. Mol. Evol. **47:** 493–500.

SMITH, N. G. C., and L. D. HURST, 1999 The causes of synonymous rate variation in the rodent genome: can substitution rates be used to estimate the sex bias in mutation rate? Genetics **152:** 661–673.

SOKAL, R. R., and F. J. ROHLF, 1995 *Biometry.* W. H. Freeman and Company, New York.

TAMURA, K., 1992 Estimation of the number of nucleotide substitutions when there are strong transition-transversion and G+C content biases. Mol. Biol. Evol. **10:** 512–526.

TAMURA, K., and M. NEI, 1993 Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. Mol. Biol. Evol. **10:** 512–526.

THOMPSON, J. D., D. G. HIGGINS and T. J. GIBSON, 1994 ClustalW—improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. **22:** 4673–4680.

WOLFE, K. H., and P. M. SHARP, 1993 Mammalian gene evolution—nucleotide sequence divergence between mouse and rat. J. Mol. Evol. **37:** 441–456.

WOLFE, K. H., P. M. SHARP and W. H. LI, 1989 Mutation rates differ among regions of the mammalian genome. Nature **337:** 283–285.

YANG, Z., 1997 PAML: a program package for phylogenetic analysis by maximum likelihood. Comput. Appl. Biosci. **13:** 555–556.

Communicating editor: G. B. GOLDING

Chapter 8. Evolution within introns

Eukaryotic protein-coding genes are divided into a number of compartments on the basis of transcription and translation. Flanking sequences surround the central portion of the gene sequence which is divided into exons which are translated and introns which are excised from the pre-mRNA before translation. In the terms of my thesis (see Chapter 1) does such gene anatomy make sense? The 5' flanking region contains the promoter region, the exons code for the gene product, the 3' flanking region contains signals for transcription termination and mRNA polyadenylation, but how does the presence of introns make evolutionary sense? I shall consider the problems of the genomic anatomy of introns at two levels: the level of the gene and the level of the nucleotide.

At the level of the gene we observe the repeated alternation of introns and exons, so the problem is one of intron copy number. The question of why eukaryotic genes contain introns at all has been hotly debated for many years (reviewed by Logsdon 1998). The introns-early (IE) hypothesis suggests that the earliest genes possessed introns, while the introns-late (IL) hypothesis proposes that genes originally arose as uninterrupted sequences of protein coding DNA into which introns subsequently inserted.

How do the alternative theories compare in terms of the roles of selection and mutation in creating current patterns of gene organisation? The IL theory suggests that introns may be present because selection is too weak to counteract their mutational spread. Introns may well be costly: speed of transcription, RNA splicing, nutrient requirements, and replication speed are all likely to be affected (Gentles and Karlin 1999; Hurst *et al.* 1996b). But if effective population sizes are too small then selection will be unable to halt the spread of introns throughout the genome. Under such a theory then, the presence of introns only make sense in terms of selection for selfish DNA: the higher the rate of horizontal transmission the better.

In contrast the IE theory proposes a possible selective benefit to introns at the level of the organism. It is argued that the presence of introns was necessary during the early evolution of protein coding genes, with groups of modular intron-exon units coming together to form genes. Because the IE theory supposes ancient introns, it implies that current patterns of introns are due to subsequent selection. Prokaryotes no longer have introns because the selection against introns is highly effective in simple organisms which consequently possess economical genomes. Eukaryotes possess introns for one of three reasons: either mutation is limiting, or introns are costly but selection is too weak, or introns may actually be selectively advantageous.

Why might introns be selectively advantageous? Conserved features may indicate stabilising selection and one clearly conserved feature of introns is the GT-AG rule of splicing junctions,. If splicing goes wrong then the protein will probably be non-functional. The need for correct splicing may lie behind the presence of compensatory selection to maintain intronic pre-mRNA structure (Kirby *et al.* 1995). However such selective arguments only explain why certain

mutations would be deleterious rather than why introns themselves are advantageous. Alternatively, the presence of introns may aid the control of alternative splicing which allows increased flexibility of expression, and which may necessitate selectively maintained pre-mRNA structures (Leicht *et al.* 1995). The existence of dsRNA editing provides a mechanism whereby intronic evolution can enable phenotypic evolution (Herbert and Rich 1999). Another possible benefit of introns may lie in error checking: if a frameshift mutation is encountered before the final exon then the mRNA will not be translated (Culbertson 1999; Nagy and Maquat 1998). Introns may act as binding sites for transcription factors (Alakokko *et al.* 1995). Finally, introns may present an advantage in the process of gene diversification. Since the cosuppression (directed mutagenesis) of repeated units is dependent on sequence similarity and requires a certain minimum homology size, the presence of introns within a duplicated protein allow for the mutagenic effects of cosuppression to be halted by rapid evolution of the introns of a gene (Kricker *et al.* 1992). Thus the presence of introns enable the maintenance of gene families which have strong protein-coding sequence homology.

So much for intron evolution at the level of the gene, but the genomic anatomy of introns can also be viewed at the level of the nucleotide. The issue now becomes one of nucleotide copy number: how does one explain intron length variation? Intron lengths are much more variable than exon lengths, which presumably means that selection on introns is weaker than selection on introns, but that does not mean that introns are completely unaffected by selection. Introns of different lengths may well carry different costs and benefits. This question if related to the C value paradox, although the fact that genes constitute such a small proportion of the eukaryotic genome means that intron length variation is unlikely to explain much genome size variation.

I have looked at intronic evolution in two ways. Firstly, I have considered the use of intronic substitution rates to infer selection. Secondly I have attempted to understand the forces responsible for the evolution of intron length in vertebrates. As with silent sites, selection on introns in mammals is likely to be weak and so mammalian intron evolution is a good place to look for the limits of selection.

Comparing intronic and synonymous substitution rates

Under neutrality both introns and synonymous sites evolve at the mutation rate, which we shall assume to be the same for two classes of site. However the equality of silent and intronic rates need not imply neutrality since the two classes of site may be under similar selective pressures. However the null hypothesis of equal rates is useful because the class of sites evolving slower is probably subject to selection, assuming that most mutations are deleterious and that mutation rates do not differ for exons and introns.

Previous comparisons of synonymous and intronic rates have indicated similar values (Li *et al.* 1985b; McVean and Hurst 1997a). I found a similar result using a much larger dataset (see Research Paper 9), and so was unable to falsify the hypothesis that both synonymous and intronic

sites are evolving neutrally in mammals. However, I did find a worrying sensitivity on intronic rate estimation to alignment procedure, which would appear to invalidate comparisons involving non-coding DNA rate estimation, as was confirmed by a subsequent methodological study (see Chapter 10 and Research Paper 11).

Intron size and base composition

The evolution of intron size and the evolution of base composition appear to be related: in humans genes in GC rich isochores tend to have short introns (Duret *et al.* 1995), and intron base composition correlates well with isochore base composition (Donofrio *et al.* 1991). Given the correlation between base composition and recombination in some mammals (Eyre-Walker 1993), two hypotheses can explain the correlation of intron length and base composition. Either recombination may induce a mutational bias towards deletions (Duret *et al.* 1995), or selection for shorter introns may be more effective in regions of high recombination (Charlesworth 1994). The discrimination of these two hypotheses is equivalent to finding whether it is mutation or selection which makes evolutionary sense of the correlation between base composition and intron size. I have analysed the relationships between intron size, base composition and recombination rates in a number of organisms (see Research Paper 10) but the discrimination of the two alternatives is not possible.

Research Paper 9. Silent sites evolve at the same rate as introns in rodents

Nick Smith and Laurence Hurst (1997)
Unpublished

Introduction

Are synonymous mutations in mammalian coding DNA neutral? The answer will provide an idea of the limits of selection in shaping genomic characters: just how finely can selection 'tune' the mammalian genome? Further, if silent sites are evolving neutrally then K_s (the synonymous substitution rate) is equal to the mutation rate (Ohta 1992). Easily aligned exonic sequence data could then be used to estimate the mutation rate and variation in K_s within the genome can be used to infer variation in mutation rate (such a rationale is adopted in McVean and Hurst 1997). The precise value of the mutation rate is in turn important for some theories of the evolution of sex (Kondrashov 1988). Furthermore if silent sites do evolve neutrally they should evolve at a constant rate, as long as mutation rate and generation times are constant.

There are several lines of evidence which suggest that most synonymous sites in mammals are effectively neutral. (1) The generation time effect is expected for neutrally evolving sites if most mutations occur in the course of gamete production, and the effect does appear to be more pronounced for synonymous than nonsynonymous sites (Li *et al.* 1996; Ohta 1995). (2) Codon usage patterns in mammals appear to be better explained by mutation patterns than selection for codon bias (Eyre-Walker 1991). (3) If synonymous mutations were selectively constrained then one would expect such deleterious mutations to reach fixation more often in regions of low recombination due to the effects of background selection and hitchhiking (Charlesworth 1994). But in rodents Y-linked genes show similar silent substitution rates to autosomal genes (McVean and Hurst 1997). (4) Processed pseudogenes which are almost certain to evolve neutrally from the time of their creation (Li *et al.* 1981) show similar substitution rates to silent sites in the few genes tested (Wolfe *et al.* 1989).

A similar method to (4) is to compare the substitution rates of intronic and silent sites. Studies based on limited sample sizes indicate that the rates are similar (Li *et al.* 1985; McVean and Hurst 1997) . If introns are evolving neutrally, as commonly thought (see below), then this result implies neutral synonymous evolution. Here we perform a similar analysis using a large number of orthologous mouse and rat genes. In addition we investigate the sensitivity of evolutionary distance estimates to alignment procedures. The measure of divergence used is the proportion of sites at which substitutions have become fixed, given by K_i and K_s for introns and

synonymous sites respectively. Hence the null hypothesis is that the mean of K_i/K_s for a large number of rodent genes does not differ significantly from one. By taking the ratio of K_i to K_s, both gene-specific (Mouchiroud *et al.* 1995) and region-specific (Sharp *et al.* 1995) effects are controlled for.

Materials And Methods

Distance Measurements

Testing the null hypothesis requires the reliable estimation of evolutionary distances. The program used was Kestim provided by Comeron who reports it to compare favourably to other modern distance estimation methods (Comeron 1995). Kestim uses a method which based on Li's (1993) method but with a putative source of bias removed. As with Li's (1993) method the Kimura two-parameter (K2P) model (Kimura 1980) is used to correct for multiple substitutions for both coding and non-coding DNA sequence data. This method is not the best for providing reliable estimates in the face of GC% variation but the inaccuracy is not anticipated to be large (see Pesole *et al.* 1995). To check the reliability of the method under GC% variation, 12 pairs of mouse-rat aligned coding sequences were analysed for K_s using both Kestim with K2P and the method of Ina (1995) which accounts for differences in GC%. The 12 genes were chosen for a wide spread of GC%, with the G+C content at four-fold degenerate sites ranging from 42% to 79%. The two methods gave very similar K_s values irrespective of GC% (a difference of less than 3% for 10 genes out of 12).

There appears no need to correct distance estimates for local GC% (as might be required if GC% influenced the local mutation rate). Silent site GC% is known to be well correlated with intron GC% and the difference between intron and silent site GC% in mammalian genes does not seem to be large (Donofrio *et al.* 1991). Furthermore rodents show a narrower isochore GC% distribution than other mammals (Mouchiroud *et al.* 1988). As a check we have obtained the relationship between four-fold degenerate site GC% and intron GC% for our data (see Results). Local GC% is no longer considered to influence mutation rates and hence K_s (for example Bernardi *et al.* 1997). In the absence of variance estimates we assume that the greater the number of compared sites in coding regions or introns the more accurate the distance estimate, hence the weighted mean of K_i/K_s will be a more reliable estimate than the unweighted mean.

Data Acquisition

A large number of data points are required in view of the high variation in K_s for rodent genes (Wolfe and Sharp 1993). Three sources were used to generate the data set considered here. The list of mouse-rat gene names in Wolfe and Sharp (1993) was scanned using NCBI Entrez to find the mouse-rat gene pairs with intronic data. The accession numbers of mouse-rat complete gene pairs were obtained from Duret *et. al.*(1995) and Ogata *et. al.*(1996). BLASTN searches

(Altschul *et al.* 1990) and intron-exon structure comparisons (using the graphical output of Entrez) were carried out to check orthologies. Orthology rather than paralogy was confirmed by looking at gene trees in HOVERGEN (Duret *et al.* 1994). The final dataset consisted of 43 gene pairs which contained a total of 136 intron pairs (see Table 3 for a list of gene names and GenBank accession numbers, further data can be obtained by emailing the corresponding author at bspns@bath.ac.uk).

Alignments

Before the evolutionary distances between orthologous sequences can be calculated the sequences must be aligned with gaps introduced to account for insertions and deletions that have occurred in the two sequences since the time of their common ancestor (page 54 in Li and Graur 1991). Generally alignment algorithms seek to find the alignment which minimises the dissimilarity or maximises the similarity between the sequences (pages 57-58 in Li and Graur 1991). This is equivalent to searching for the alignment which accounts for the minimum amount of evolution. So the process of alignment, a procedure which is required before evolutionary distances can be measured, itself makes assumptions about molecular evolution. In addition for an alignment to be scored various parameters must be set which reflect the expected frequency of matches, mismatches, and gaps. The quality of alignments will depend on the values of these parameters, for example the best mismatch scores to use will depend upon whether one is seeking relatively diverged or closely related sequences (Altschul 1997). There exists no general theory to guide the selection of biologically relevant gap scores (Altschul 1997).

Despite the fundamental problems of alignment outlined above, the alignment of exons in this study was robustly insensitive to scoring parameters. In uncertain cases the alignment was repeated using protein data. Furthermore alignments could be adjusted by hand to remove gaps of lengths not multiples of three (equivalent to assuming that frameshifts never reach fixation). Hence only one set of K_s values was obtained.

The problems of alignment became more pronounced when analysing the intronic data. Orthologous introns often had very different lengths and were highly diverged which meant that the alignment program had a much greater number of possible alignments to choose between. Significant sensitivity to scoring parameters was observed. The alignment program used in this study was the PILEUP program of the Wisconsin Package provided by the Genetics Computer Group (GCG) which offers two gap penalty values to be altered by the user: gap creation penalty and gap extension penalty. These settings are important to the present study because different alignments give different distance estimates. Altering the gap penalty parameters means that a different alignment is optimal. If gap penalties are low, then the evolutionary distance estimate will also be low (in the extreme if there are no gap penalties then the alignment will consist of many gaps which separate runs of near-perfect matches).

Scoring Alternative Alignments

We have attempted five different methods to score alternative parameter settings. These methods should allow us to say which parameter setting of those test is the 'best' (being the truest representation of evolutionary history).

Method 1: comparing alignments at the conserved ends of introns

The ends of introns have conserved sequences, including six nucleotides at the 5' end and three nucleotides at the 3' end (page 914 in Lewin 1994) (these sequences were removed from intron alignments before distance estimation, just as stop and start are ignored in the calculation of K_s). One would expect any gaps that appear in these strongly conserved regions to be more selectively constrained than gaps appearing elsewhere. Furthermore a 'good' alignment should match the conserved ends of introns because we can be sure that these ends should be aligned. With PILEUP it was possible to use these predictions since terminal gaps are allowed. The introns were aligned using different gap scores. One set of alignments used the default penalty values of 5.0 gap creation and 0.3 gap extension (*default* set), while for a second the default values were halved (*easy* set), and for a third both gap values were set to zero (*zero* set). These penalty values are relative to the match and mismatch penalty scores of 0 and 1 respectively (PILEUP seeks the alignment with the lowest penalty score). The three sets of alignments were scored with one mark for each intron end alignment without gaps, and no mark if a gap was present. A 'better' set of alignments might be predicted to have a higher score than a 'worse' one, but the 'best' alignment would not be expected to score full marks.

The *default* set scored 75.0%, the *easy* set 84.3%, and the *zero* set 85.7%. Despite the *zero* set having the highest score, several of its alignments appeared clearly incorrect since enormous numbers of gaps were introduced when the two sequences to be aligned differed considerably in length (rather than just one or two large gaps which would seem a more biologically valid interpretation). Since the *zero* penalty protocol produced more gaps (see method 2 below) it is perhaps surprising that the *zero* set had fewer gaps in the conserved ends. But the *zero* set penalty scores mean that matches are strongly favoured. Since the ends of introns are well conserved (indeed this is the basis of the test) it is not surprising that an alignment which maximises matches causes the ends to be aligned without gaps. Therefore one would not necessarily expect the highest score for the alignment best reflecting evolutionary history, and hence the test is likely to be flawed.

Method 2: comparing alignments to a mismatches:gaps reference

Ogata et. al. (1996) took orthologous primate non-coding sequences and using a three species strategy produced over 6 knt. of alignments which were insensitive to the scoring parameters of the multiple alignment programs CLUSTAL-V. These alignments contained 3.8% mismatches (indicating substitutions if the alignments were correct) and 2.5% gaps, a ratio of ~1.5:1 mismatches:gaps. If one takes this ratio as an indication of mammalian molecular evolution processes, then it is possible to use it as a benchmark against which to test the three different mouse-rat alignment sets.

This method makes a fair number of unjustified assumptions. There may be differences between rodent and primate mutational patterns, and between introns and other non-coding sequences. Also the larger population sizes of rodents may cause differences in relative strengths of selection and drift in the two groups. One indication of primate-rodent differences is the statistically significant tendency of human introns to be larger than those of rodents (Duret, Mouchiroud and Gautier 1995; Ogata, Fujibuchi and Kanehisa 1996). Another is the fact that deletions in processed pseudogenes accumulate faster in rodents than in humans (Graur *et al.* 1989). Finally, the total amount of gaps and mismatches may differ between the two groups, in which case the group with the larger total might have been exposed to more selection.

The three alignment sets described in method 1 were analysed for total alignment length, %mismatches, %gaps, and the ratio of mismatches:gaps (see Table 1). Since the largest gap recorded in the Ogata group's data was 16 nt. (Ogata, Fujibuchi and Kanehisa 1996), all long gaps (>50 nt) were removed from consideration as were those introns which returned a similarity score of >200 (following a protocol set out in Ogata, Fujibuchi and Kanehisa 1996) when checked against a repetitive element database using BLASTN (Altschul *et al.* 1990). As the gap penalties decrease the proportion of mismatches decreases, the proportion of gaps increases, and the total alignment length increases. All these changes are in the directions expected from a consideration of alignment algorithms, and lead to the ratio of mismatches:gaps decreasing as the gap penalties decrease.

The reference ratio turns out to be intermediate between the *default* and *zero* values, and close to that of the *easy* alignment protocol. Therefore this test would rate the *easy* alignment protocol over both the *default* and *zero* protocols.

Method 3: comparing alignments to an average gap length reference

Gu and Li (1995) have worked out the size distribution of 78 human processed pseudogenes for insertions, deletions, and indels. Checking the size distribution of indels for various gap score settings against this benchmark allows the relative gap creation and extension penalty sizes to be evaluated. To simplify the analysis Gu and Li's indel distribution data was used to calculate an average gap length of 3.53 nt. One problem with this analysis is that while processed pseudogenes are very likely to evolve neutrally, such an assumption clearly cannot be extended to introns in a study which investigates how introns evolve. In addition the same problems discussed in method 2 with regard to possible primate-rodent differences will apply, though the frequencies of a single nucleotide gap are similar for human and rodent processed pseudogenes (Gu and Li 1995). If introns are constrained selectively then would one expect the average gap length to be less than or greater than the processed pseudogene value? There seems little reason to advocate either an increase or decrease in gap length. Deletions are more common than insertions in sequence evolution (Gu and Li 1995), but that still leaves the problem of whether large or small deletions would be more heavily constrained.

Analysis of the alignment sets produced the following results: the *default* set had an average gap length of 5.81, while the *easy* set gave 5.46, and the *zero* set 1.80. These data suggest that the *easy* protocol is better than the *default* one, while this test cannot discriminate between the *easy* and *zero* protocols.

Method 4: comparing multiple and pairwise alignments

A multiple alignment using more than two orthologous sequences should be more reliable (less penalty gap sensitive) than a pairwise alignment. Using multiple sequences of common ancestry means that there is more information revealing evolutionary history, but on the other hand multiple alignment programs need to be more complex and make more assumptions than pairwise alignment programs. Ogata *et.al.* (1996) report extreme scoring parameter insensitivity in their multiple alignments, while Gu and Li (1995) had to remove "some pseudogenes" from their analysis due to gap penalty sensitivity. Multiple alignments could not have been used as the core of this study because the number of genes would have been greatly reduced.

A search was undertaken for mammalian intron-containing orthologues to the rat-mouse genes, using NCBI Entrez and BLAST (Altschul *et al.* 1990) searches. Many human orthologues were found but, considering that human introns are significantly larger than rodent ones (Ogata, Fujibuchi and Kanehisa 1996), only multiple alignments with mouse, rat, and at least one non-human other mammal were considered. This gave 7 genes with a total of 15 introns. The multiple alignments were carried out using PILEUP with the same three parameter sets as for the pairwise alignments. The sensitivities of the pairwise and multiple alignments to penalty gap settings were compared by summing for all introns the square of divergences from the average K_i of the three parameter settings. This revealed that the multiple alignments were more variable than the pairwise alignments as the penalty gap scores changed. This method is hence uninformative.

Method 5: partitioning the data according to penalty gap sensitivity

All the pairwise alignments were compared, and the genes were ranked according to penalty gap sensitivity (again measured using the sum of the squares of divergences from the average). On this basis, the genes were divided into two groups: the 21 least sensitive genes and the 22 most sensitive genes. These two groups are compared in the results section below.

Results

Comparing Intron And Synonymous Rates

The three sets of intronic alignments were analysed for K_i and L_i (the number of intronic sites compared), and the one set of exonic alignments was analysed for K_s and L_s (the number of synonymous sites compared). Each gene was weighted according to the formula $(L_s.L_i)/(L_s+L_i)$ in order to produce a weighted K_i/K_s mean. For the three sets of alignments a plot of K_i/K_s (y-axis)

against weight (x-axis) for all 43 genes showed that, as expected, the variation in K_i/K_s decreased as gene length increased. The linear regression plots had gradients not significantly different from *zero* in all three cases, which implies that Ki/Ks does not vary systematically with weighting.

For each alignment protocol we calculated the weighted means and weighted standard errors of K_i/K_s. The *default* set gave a weighted K_i/K_s of 1.34 +/- 0.15 (mean +/- S.E.), for the *easy* set K_i/K_s was 1.07 +/- 0.08, and for the *zero* set 0.63 +/- 0.05 (see Table 2). The normal distribution was used to determine the 95% confidence intervals (although neither Ks nor Ki can be assumed to distributed normally ratios do tend to be normally distributed). The null hypothesis predicts a K_i/K_s of 1, and so both the *default* and *zero* data sets reject the null hypothesis while the *easy* data set fails to reject (see Figure 1).

The Methods section details the results of three ways of evaluating the alignment sets. The test of aligned conserved sequences (though likely to be flawed) gave *easy* better than *default*. The test of mismatches:gaps gave *easy* as better than both *default* and *zero*. The test of average gap length gave *easy* better than *default*. So assuming that the tests are reasonable, we have good evidence for *easy* being better than *default*, but weaker evidence for *easy* being better than *zero*. In a similar study to ours Li *et al.* (1985) obtained nucleotide substitution rates of 3.70 and 4.65 (both values in units of 10^{-9} substitutions per site per year) for introns and synonymous sites respectively, from an analysis of more than ten mammalian genes. This result result fits in with our finding that it is easier to reject introns evolving faster than silent sites than to reject silent sites evolving faster than introns. As a final test, the genes were divided into two groups according to the sensitivity of K_i to penalty gap changes (in effect the same test as performed by Gu and Li (1995) to remove dubious multiple alignments). Again the weighted means and S.E.s were calculated.

This method assumes that those alignments which are less sensitive to penalty gap parameters are more reliable indicators of K_i. If the 21 least sensitive genes are considered then the weighted K_i/K_s means become more bunched although the differences between the alignment protocols remain significant, and would appear to be converging towards the *easy* value of around one. Furthermore the 22 most sensitive genes give more wide-ranging weighted K_i/K_s means. In terms of which alignment protocols now reject the null hypothesis, the division of the data causes very little change. That the most variable genes show no qualitative difference from the least variable suggests this method might be unsatisfactory, but with smaller numbers of data points the errors are bound to increase.

That the K_i/K_s means of the genes which show low penaly gap sensitivity exhibit low sensitivity to penalty gap values themselves comes as no surprise. But the K_i/K_s value that the different protocols approach when the more reliable genes are tested does suggest a weighted K_i/K_s mean close to 1. Hence this test fails to reject the null hypothesis.

Evidence For Gene-Specific Mutation Rates

Assuming that both K_i and K_s are unbiased measures of the mutation rate then we can ask whether there occur systematic differences in the mutation rates between genes. If there were no gene-specific differences in the mutation rate then K_i and K_s should not covary. The highest correlation was 0.577 ($p<<0.01$) for the *easy* set, followed by 0.479 ($p<0.01$) and 0.201 ($p>0.05$) for the *default* and *zero* sets respectively. Since we consider the *easy* alignment protocol to be the 'best', this we consider to be strong evidence for gene-specific differences and consistent with previous reports (Ogata, Fujibuchi and Kanehisa 1996). Variation in GC content does not explain variation in weighted K, and hence the similar compositions of silent sites and introns (see below) do not cause the K_i-K_s correlation.

It has been established that X-linked genes have lower Ks values than autosomal genes (McVean and Hurst 1997). To what extent does this effect account for the result above? If only those genes which are known to be located on autosomes are considered, the K_i-K_s correlation for the *easy* alignment protocol remains significant (0.545, $p<0.01$). Hence the observed correlation provides evidence that there are gene-specific differences in mutational rates between autosomal genes.

GC Content In Introns And Silent Sites

If mutational pressures alone determine both silent site and intronic composition, then GC_4 (G+C content at four-fold degenerate silent sites) and GC_{intron} (intronic G+C content) will be equal if the mutation patterns of neighbouring silent sites and introns are the same (as seems likely). However we found that GC_4 is significantly higher than GC_{intron}. The GC4%s of mouse and rat were not significantly different (Wilcoxon 1-sample median and 95% confidence limits of mouse-rat were -0.5 and -3.0 to 1.2) and so mouse-rat mean values were used. A Wilcoxon 1-sample test of GC_4-GC_{intron} gave a median value of 14% and 95% confidence limits of 12 and 17%. In only one gene out of 43 was $GC_{intron} > GC_4$. A linear regression plot of GC_4 against GC_{intron} (Figure 2) shows that GC_4 values start about 10% higher than GC_{intron} and this difference becomes greater as GC content increases. These findings fit in with previous comparisons such as GC_3 (G+C content at third codon positions) versus GC_{intron} in humans (Clay *et al.* 1996).

Discussion

K_s and K_i are similar

We have shown (1) that estimates of intronic substitution rates are significantly sensitive to alterations of alignment protocol, (2) that the alignment protocol we consider best

indicates that K_i is not significantly different from K_s, (3) that under the same protocol there is strong evidence for systematic differences between genes in their mutation rates (not explicable by compositional variation) and (4) that GC content of four-fold degenerate sites is significantly higher than that of neighbouring introns.

At first sight findings (2) and (4) appear contradictory. K_s equal to K_i is consistent with neutral evolution, yet the difference in GC content between introns and silent sites suggests that selection operates differently in the two classes of sequence. How can we unify these two findings? One possibility is that selection does not act on individual mutations in silent sites or introns but instead acts on modifiers of GC composition. Since a modifier can affect many sites the selection on a modifier can be strong enough to overcome drift even when selection at individual silent sites is too weak to cause compositional changes. Hence the systematic compositional difference does not constitute firm evidence for selection on individual silent sites.

An alternative explanation would be that selection operates on both classes of site to a similar extent (hence causing similar slow-downs of the evolutionary rate) but that selection acts on composition in different ways. The question then is whether both introns and silent evolve neutrally.

Do both introns and silent sites evolve neutrally?

Our failure to refute the null hypothesis cannot on its own be interpreted as evidence for the neutral theory since the null hypothesis would also follow from introns and synonymous sites undergoing similar levels of selection. Is it likely that selection would favour point substitutions in either class of site?

It is unreasonable to assume that in an infinity large population selection will be unable to influence the composition of silent sites. Indeed there exists abundant evidence that silent sites are affected by selection in groups other than mammals. Numerous selective mechanisms can be hypothesised: RNA and DNA structural demands, protein folding (Adzhubei *et al.* 1996), selection for G+C content (Bernardi 1995), and translational accuracy and speed (see Akashi 1994). There exist correlations between codon usage patterns, tRNA abundances and binding preferences, gene expression levels, synonymous substitution rates, and gene functional importance in organisms as diverse as *E.coli*, *Saccharomyces cerevisiae*, and *Drosophila melanogaster* (references in Sharp 1989). This apparent co-adaptation provides strong evidence of selection.

Similarly selection could affect intronic substitutions. Although the evolution of introns has generally been assumed to be neutral (e.g. intronic substitution rate used as estimate of mutation rate in Li *et al.* 1996), a variety of selective mechanisms on introns appear plausible (minimum size, regulatory elements and high free energy proposed in Leicht *et al.* 1995). To the best of our knowledge selection on point mutations in introns has not been demonstrated in any group of organisms let alone mammals.

The explanation for the effects of selection being weaker in mammals is the difference in population sizes (Li 1987). In general if the selective coefficient of an allele is less than the reciprocal of the long-term evolutionary effective population size, then the behaviour of that allele will be no different to that of a neutral allele (Kimura 1983). Mammalian species are thought to have much lower effective population sizes than for example *Drosophila* species, which will make drift and mutation more potent mediators of codon usage change relative to selection. For example, effective population sizes in rodents are likely to be one or two orders of magnitude smaller than those of most *Drosophila* populations (Nei and Graur 1984).

In this study selection will only be revealed if it affects the majority of either intronic or silent sites, and even then only if the stabilising selection is strong enough to cause a significant reduction in K_s or K_i. Whilst the first requirement may be fulfilled, the second is unlikely to be owing to the relatively small population sizes of mammals. It should be noted however that there is evidence for selective constraints on silent substitutions in calmodulin (Britten 1993) and for conserved introns in procollagen (Alakokko *et al.* 1995). Such examples are probably exceptional and should not affect our results.

An estimate of the mutation rate

Since K_i and K_s are similar we can conclude that silent site substitution rates will provide as (in)accurate an estimate of the mutation rate as intronic substitution rates. Given that exonic alignment is considerably less susceptible to gap penalty settings, this conclusion has practical benefits. For the reasons given above (Discussion B) this may be assumed to be a reasonable estimate of the mutation rate.

The mean weighted K is 0.186, 30% higher than previously reported on the basis of limited intronic data (Li *et al.* 1996). The Li et al. study (Li *et al.* 1996) used a slightly different distance estimation method from that used here, but such differences would not be expected to cause such a large change (Comeron 1995). Estimates of the time of the mouse-rat divergence vary from 10 to over 30 million years (discussed in Li *et al.* 1996). These yield mutation rates of between 9.3 and 3.1×10^{-9} per base pair per year.

Acknowledgements

Thanks to: Josep Comeron for Kestim; all the authors of the papers from which data were obtained; HGMP and Cameron Dunn for computing resources; Gil McVean for many helpful comments; and all those at the Centre for Mathematical Biology.

Figures

Figure 1: The means and 95% confidence intervals of the weighted ratio of intronic substitution rate to synonymous substitution rate (Ki/Ks), with data for different intron alignment protocols (from default to zero gap penalties are reduced) and for different data subsets (classified on the basis of substitution rate sensitivity to intron alignment protocol).

Figure 2: A linear regression plot of the GC content of four-fold degenerate sites (GC_4) against the GC content of introns (GC_{intron}). For comparison a line of slope 1 and intercept at the origin ($GC_4 = GC_{intron}$) is also shown.

Table 1: Results for alignment comparison method 2: the default, easy and zero intron alignment protocols (gap penalties reduced from default to zero) are compared against reference values (from Ogata *et al.* 1996) of proportion of alignment consisting of mismatches, proportion of alignment consisting of gaps, ratio of mismatches to gaps, and total alignment length.

Table 2: Statistics for the weighted ratio of intronic substitution rate to synonymous substitution rate (Ki/Ks), with data for different intron alignment protocols (from default to zero gap penalties are reduced) and for different data subsets (classified on the basis of substitution rate sensitivity to intron alignment protocol).

Table 3: Details of the 43 genes used in this study, giving gene name and GenBank accession numbers of the mouse and rat orthologues.

Figure 1

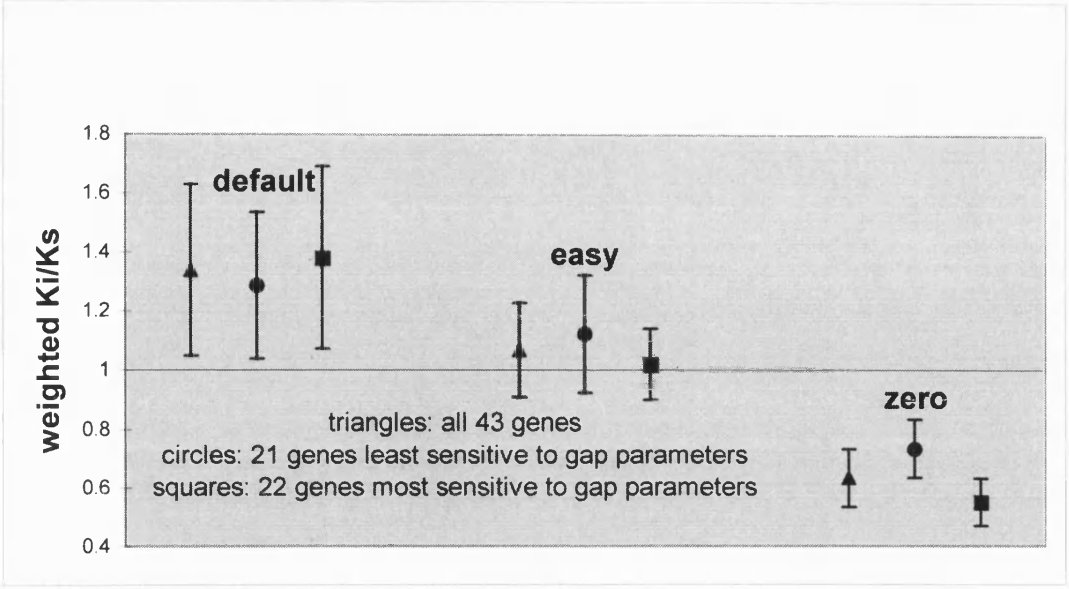


Figure 2

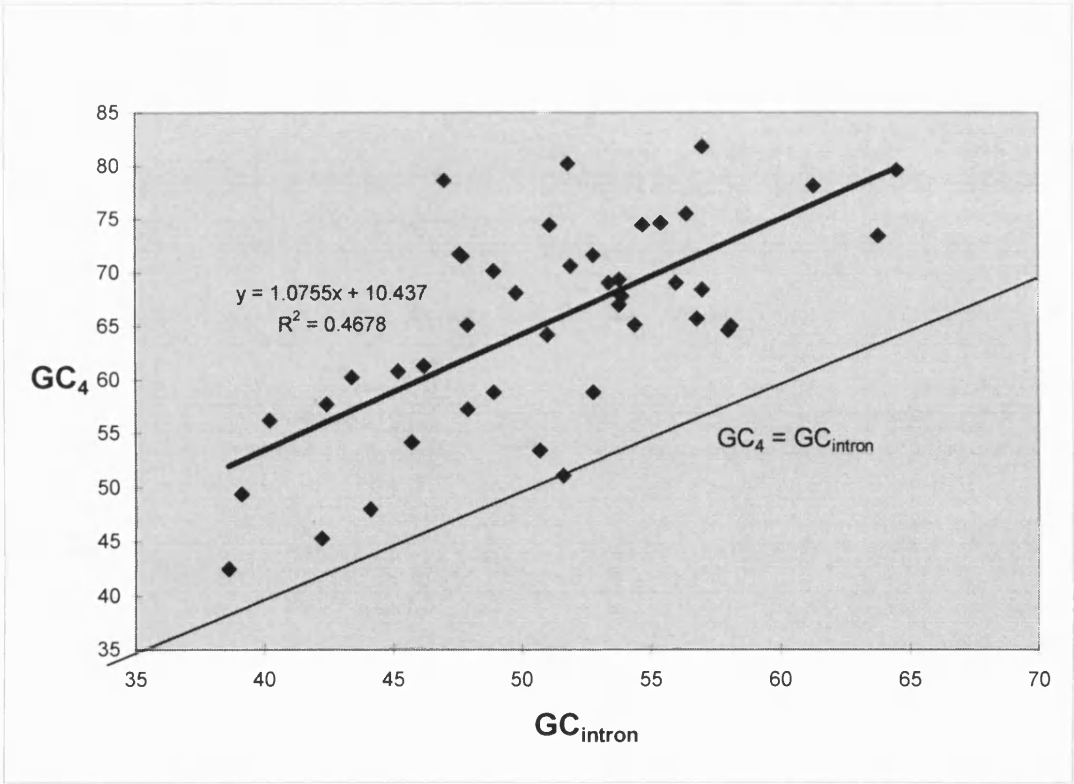


Table 1

alignment	%m	%g	m/g	length (nt)
default	16.1	8.8	1.84	13494
easy	14.6	9.8	1.5	13535
zero	8.8	17.5	0.5	13754
Ogata *et. al.*	3.8	2.5	1.52	>6000

Table 2

alignment	group of genes	Ki/Ks mean	S.E.	95% intervals
default	all 43	1.34	0.15	0.29
	least sensitive 21	1.29	0.13	0.25
	most sensitive 22	1.38	0.16	0.31
easy	all 43	1.07	0.08	0.16
	least sensitive 21	1.12	0.1	0.2
	most sensitive 22	1.02	0.06	0.12
zero	all 43	0.63	0.05	0.1
	least sensitive 21	0.73	0.05	0.1
	most sensitive 22	0.55	0.04	0.08

Table 3

gene name	mouse	rat
alpha-1-acid glycoprotein	m17376	m10614
skeletal alpha-actin	m12347	j00692
adenine phosphoribosyltransferase	m11310	l04970
alpha-lactalbumin	m87863	x00461
anti-mullerian hormone	x63240	s98336
apolipoprotein A-I	x64263	j02597
apolipoprotein A-IV	m13966	m13508
apolipoprotein E	d00466	j02582
osteocalcin	l24429	m23637
c-myc proto-oncogene	j00374	y00396
creatine kinase B	m74149	m18668
alpha-B-crystallin	m73741	u04320
gamma-A-crystallin	k02587	m19359
gamma-E-crystallin	x57855	m19359
cystatin beta	u59807	d10607
elastase II	x04576	j00731
gonadotrophin-releasing hormone	m14872	m31670
insulin 2	x04724	v01243
immediate-early serum-responsive je	m19681	x17053
mGK-1 kallikrein	v00829	m19647
lutenising hormone beta	u25145	j00749
metallothionein 2	k02236	m11794
myogenic differentiation 1	x61655	m84176
atrial natriuretic polypeptide	k02781	k02062
type B natriuretic peptide	s58667	m60731
n-myc proto-oncogene	x03919	x63281

Table3 continued

oxytoxin-neurophysin	m88355	k01701
cytochrome P450c(1)	m10021	m26129
cytochrome P450d(3)	m10022	k03241
reg III beta	d63360	l07127
peripherin	x59840	m26232
phenylethanolamine N-methylase	l12687	x75333
ornithine decarboxylase	x06572	j04792
regenerating protein 1	d14010	d26164
secretin	u07568	m64033
preprosomatostatin	x51468	j00787
surfactant protein SP-C	m38314	u07796
thy-1 antigen	x03151	x03152
thymidine kinase	m68489	x54173
tnf alpha	y00467	d00475
tnf beta lymphotoxin	y00137	l00981
trypsin Ta	x04577	j00778
thyrotropin beta-subunit	m22739	m14499
uncoupling protein	u63418	x12925

References

ADZHUBEI, A. A., I. A. ADZHUBEI, I. A. KRASHENINNIKOV and S. NEIDLE. 1996. Non-random usage of 'degenerate' codons is related to protein three-dimensional structure. FEBS Letters 399: 78-82.

AKASHI, H. 1994. Synonymous codon usage in *Drosophila melanogaster* - natural selection and translational accuracy. Genetics 136: 927-935.

ALAKOKKO, L., A. P. KVIST, M. METSARANTA, K. I. KIVIRIKKO, B. DECROMBRUGGHE, D. J. PROCKOP and E. VUORIO. 1995. Conservation of the sizes of 53 introns and over 100 intronic sequences for the binding of common transcription factors in the human and mouse genes for type-Ii procollagen (Col2a1). Biochemical Journal 308: 923-929.

ALTSCHUL, S. F. 1997. Sequence comparison and alignment. Pp. 137-168 *in* M. J. BISHOP, AND C. J. RAWLINGS, eds. DNA and protein sequence analysis. Oxford University Press, Oxford.

ALTSCHUL, S. F., W. GISH, W. MILLER, E. W. MYERS and D. J. LIPMAN. 1990. Basic local alignment search tool. J. Mol. Biol. 215: 403-410.

BERNARDI, G. 1995. The human genome: Organization and evolutionary history. Annu. Rev. Genet. 29: 445-476.

BERNARDI, G., D. MOUCHIROUD and C. GAUTIER. 1997. Isochores and synonymous substitutions in mammalian genes. Pp. 137-168 *in* M. J. BISHOP AND C. J. RAWLINGS, eds. DNA and Protein sequence analysis. Oxford University Press, Oxford.

BRITTEN, R. J. 1993. Forbidden Synonymous Substitutions in Coding Regions. Mol. Biol. Evol. 10: 205-220.

CHARLESWORTH, B. 1994. The effect of background selection against deleterious mutations on weakly selected, linked variants. Genet. Res. 63: 213-227.

CLAY, O., S. CACCIO, S. ZOUBAK, D. MOUCHIROUD and G. BERNARDI. 1996. Human coding and noncoding DNA - compositional correlations. Mol. Phylog. Evol. 5: 2-12.

COMERON, J. M. 1995. A method for estimating the numbers of synonymous and nonsynonymous substitutions per site. J. Mol. Evol. 41: 1152-1159.

DONOFRIO, G., D. MOUCHIROUD, B. AISSANI, C. GAUTIER and G. BERNARDI. 1991. Correlations between the compositional properties of human genes, codon usage, and amino-acid-composition of proteins. J. Mol. Evol. 32: 504-510.

DURET, L., D. MOUCHIROUD and C. GAUTIER. 1995. Statistical-analysis of vertebrate sequences reveals that long genes are scarce in GC-rich isochores. J. Mol. Evol. 40: 308-317.

DURET, L., D. MOUCHIROUD and M. GOUY. 1994. Hovergen - a database of homologous vertebrate genes. Nucleic Acids Res. 22: 2360-2365.

EYRE-WALKER, A. 1991. An analysis of codon usage in mammals: selection or mutation bias? J. Mol. Evol. 33: 442-449.

GCG. Program Manual for the Wisconsin Package,Version 8, September 1994, Genetics Computer Group, 575 Science Drive, Madison, Wisconsin, USA 53711.

GRAUR, D., Y. SHUALI and W. H. LI. 1989. Deletions in processed pseudogenes accumulate faster in rodents than in humans. J. Mol. Evol. 28: 279-285.

GU, X., and W. H. LI. 1995. The size distribution of insertions and deletions in human and rodent pseudogenes suggests the logarithmic gap penalty for sequence alignment. J. Mol. Evol. 40: 464-473.

KIMURA, M. 1980. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. J. Mol. Evol. 16: 111-120.

KIMURA, M. 1981. Possibility of extensive neutral evolution under stabilising selection with special reference to nonrandom usage of synonymous codons. Proc. Natl. Acad. Sci. USA 78: 5773-5777.

KIMURA, M. 1983. The Neutral Theory of Evolution. Cambridge University Press, Cambridge.

KONDRASHOV, A. 1988. Deleterious mutations and the evolution of sexual reproduction. Nature 336: 435-440.

LEICHT, B. G., S. V. MUSE, M. HANCZYE and G. CLARK. 1995. Constraints on intron evolution in the gene encoding the myosin alkali light-chain in *Drosophila*. Genetics 139: 299-308.

LEWIN, B. 1994 Genes V. Oxford University Press, Oxford.

LI, W. H. 1987. Models of nearly neutral mutations with particular implications for nonrandom usage of synonymous codons. J. Mol. Evol. 24: 337-345.

LI, W. H. 1993. Unbiased estimation of the rates of synonymous and nonsynonymous substitution. J. Mol. Evol. 36: 96-99.

LI, W. H., C.-I. WU and C.-C. LUO. 1985. Evolution of DNA sequences. Pp. 1-94 *in* R.J.MacIntyre, ed. Molecular Evolutionary Genetics. Plenum, New York.

LI, W. H., D. L. ELLSWORTH, J. KRUSHKAL, B. H. J. CHANG and D. HEWETTEMMETT. 1996. Rates of nucleotide substitution in primates and rodents and the generation time effect hypothesis. Mol. Phylog. Evol. 5: 182-187.

LI, W. H., T. GOJOBORI and M. NEI. 1981. Pseudogenes as a paradigm of neutral evolution. Nature 292: 237-239.

LI, W.-H., and D. GRAUR. 1991 Fundamentals of molecular evolution. Sinauer associates, Sunderland, Massachusetts.

MCVEAN, G. T., and L. D. HURST. 1997. Evidence for a selectively favourable reduction in the mutation rate of the X chromosome. Nature 386: 388-392.

MOUCHIROUD, D., C. GAUTIER and G. BERNARDI. 1988. The compositional distribution of coding sequences and dna-molecules in humans and murids. J. Mol. Evol. 27: 311-320.

MOUCHIROUD, D., C. GAUTIER and G. BERNARDI. 1995. Frequencies of synonymous substitutions in mammals are gene-specific and correlated with frequencies of nonsynonymous substitutions. J. Mol. Evol. 40: 107-113.

NEI, M., and D. GRAUR. 1984. Extent of protein polymorphism and the neutral mutation theory. Evol. Biol. 17: 73-118.

OGATA, H., W. FUJIBUCHI and M. KANEHISA. 1996. The size differences among mammalian introns are due to the accumulation of small deletions. FEBS Letters 390: 99-103.

OHTA, T. 1992. The nearly neutral theory of molecular evolution. Annu. Rev. Ecol. Syst. 23: 263-286.

OHTA, T. 1995. Synonymous and nonsynonymous substitutions in mammalian genes and the nearly neutral theory. J. Mol. Evol. 40: 56-63.

PESOLE, G., G. DELLISANTI, G. PREPARATA, and C. SACCONE, . 1995. The importance of base composition in the correct assessment of genetic-distance. J. Mol. Evol. 41: 1124-1127.

SHARP, P. M. 1989.Evolution at 'silent' sites in DNA. Pp. 24-32 *in* W. G. HILL, and T. F. C. MACKAY, eds. Evolution and animal breeding; reviews on molecular and quantitative approaches in honour of Alan Robertson. CAB International, Wallingford, U.K.

SHARP, P. M., M. AVEROF, A. T. LLOYD, G. MATASSI and J. F. PEDEN. 1995. DNA-sequence evolution - the sounds of silence. Philos. Trans. R. Soc. B. 349: 241-247.

WOLFE, K. H., and P. M. SHARP. 1993. Mammalian gene evolution - nucleotide-sequence divergence between mouse and rat. J. Mol. Evol. 37: 441-456.

WOLFE, K. H., P. M. SHARP and W. H. LI. 1989. Mutation rates differ among regions of the mammalian genome. Nature 337: 283-285.

Research Paper 10. Small introns tend to occur in GC rich regions in some but not all vertebrates

Laurence Hurst, Clair Brunton and Nick Smith (1999)
Trends Genet. **15** 437-439.

There exists considerable variation in the size of introns both within and between species. It has been reported that, for some mammals and birds, genes in GC rich isochores may be both shorter (total intron size + total exon size) and more compact (total intron size/total exon size) than genes in isochores of lower GC content[1]. Does this mean that the introns are shorter in GC rich regions and is this generally true within the vertebrates? To address these issues we have analysed the covariance of intron size and local GC composition in a mammal, a bird, a fish and an amphibian.

A covariance of intron size with GC composition would be potentially informative of some of the forces affecting intronic dimensions as GC% and recombination rate are known to covary within mammals[2]. Indeed, it has been hypothesized that if recombination induces deletions introns might be smaller in GC rich regions owing to a mutational bias[1]. However, an alternative selectionist model can also be imagined. If longer introns are slightly deleterious then, as selection is more efficient when the local recombination rate is high[3], small deletions and insertions are more likely to be "seen" by selection in GC rich regions. Here we additionally investigate this putative link between recombination rates and GC content/intron sizes.

Accession numbers were obtained from the FTP site as indicated in Duret *et al.*[1] and from Hovergen[4]. Extraction of data from each database entry was automated using information in the Genbank accession files. This allowed us to calculate the size of each intron. Partial intron sequences were not included. The complete exons for each gene were also extracted and a GC3 percentage calculated. The raw data on intron sizes is skewed with a tail towards larger intronic dimensions. This was corrected by Log transformation.

Within humans we find a significant negative correlation between GC content and intron size (Log (Intron size)=-0.49 GC3 + 2.93; P value on slope >0.0001, N=1211 introns; see Fig. 1). This is consistent with the hypothesis that introns are smaller where the recombination rate is highest. Analysis of mouse (Duret pers. comm.) and rat data ($P<0.002$) indicate the same tendency. Sample size limitations prevent firm conclusions being drawn in other mammals (see below).

However, it must be noted that, while the statistics are very highly significant, there is also very considerable residual variance in intron size that is not explained by GC content (see Fig 1).

Data from outside of the mammals is suggestive of a potential difference between cold and warm blooded species, as there might be as regards isochores[5]. Chickens, like mammals, have significantly larger introns in AT rich isochores, (Log (Intron size)= -0.67 GC3% +2.88; P value on slope >0.0001, N=313 introns). In *Xenopus* there is a significant positive correlation of GC3 percentage and intron size, indicating the opposite pattern (Log (Intron size)= 2.28 GC3% +1.51; P value on slope >0.0001, N=83 introns). In *Fugu* the correlation is also positive and near significance (Log (Intron size)=0.5 GC3% +1.93; P value on slope =0.09, N=260 introns).

The possible warm blooded/cold blooded dichotomy needs further elucidation, most especially as the sample size for *Xenopus* is considerably lower than that for *Fugu*, but it is the former that provides a significant statistic. It must also be noted that the investigation of comparative introns sizes is heavily influenced by ascertainment biases, as small introns tend to be fully sequenced earlier and more often than large ones. So, for example, in nine species of mammals that we have examined, the best predictor of mean intron size in a species is the number of introns sequenced (Spearman Rank correlation of mean intron size per species versus number of introns sampled per species, $P<0.01$). Likewise, our analysis (data not shown) of orthologous mouse-rat introns, finds the mean size of these to be around two thirds that of an order of magnitude larger non-orthologous set. This is to be expected if small introns are sequenced first, as the probability that the same intron has been sequenced in two species is approximately a function of the square of the probability that it has been sequenced in one. Given the possibility of ascertainment problems, we are cautious about the *Xenopus* result, but given that the same pattern is seen in *Fugu*, we consider the finding worth reporting.

Considering these difficulties and given that the GC content of *Xenopus* and human genes are correlated[5], it is then worthwhile asking whether the size of orthologous introns in *Xenopus* and humans are also correlated and how this might relate to the above result. We examined Hovergen to identify orthologous *Xenopus* human genes and found a total of 33 orthologous introns. Unfortunately this sample size is too limited to make many firm conclusions. Importantly, in this sample, GC3% and Log intron size of the *Xenopus* genes do not correlate, so the sample may not be representative. We can, however, report that, while on average human and *Xenopus* introns are no

different in size (mean of Human intron /*Xenopus* intron = 1.04), there is very considerable variation (standard deviation of Human intron /*Xenopus* intron = 1.21). Some *Xenopus* introns are over ten times larger than the human ortholog while some human ones are nearly five times larger than the *Xenopus* ortholog. There is, at best, only a weak tendency for the intron sizes to correlate (Log (Human intron size)=1.46+0.378 Log (Xenopus intron size), P=0.1). We do not find any significant relationship between the GC3 content of the flanking exons in *Xenopus* and the size ratio of the orthogs. The only strong conclusion is that with the amount of variation in size between the orthogs, there clearly has been plenty of size evolution of the introns.

Given the diversity of GC3% versus intron size results, one might suspect that intron size, GC3% and recombination rate are not generally causally related. Indeed, whilst three way alignment of orthologous mammalian introns indicates a mutational bias in favour of deletions, no covariance of GC content and size difference between introns could be found, although intronic size difference and local mutation rate do covary[6].

An alternative possibility is that recombination, GC% and intron size do covary, but not in the same way in all vertebrates. It might be that in chickens and mammals there is more recombination in GC rich regions, but in the cold-blooded species the opposite pattern is found. We are unaware of pertinent data to test the prediction as regards *Xenopus* and *Fugu* (or indeed any other cold-blooded vertebrate). However, we can provide a test of the hypothesis as regards chickens. If GC% and recombination rate were related in the above manner, then the Z chromosome should have a GC content lower than that of autosomal genes, as the former recombines in males alone (except in the pseudoautosomal region). The W chromosome should have an even lower figure as it never recombines.

Chickmap (www.ri.bbsrc.ac.uk/chickmap/) provides information on map position of chicken genes. From here we have derived a list of genes whose full cDNA was known and that were also known to be either Z-linked (N=7) or autosomal (N=23). We find that the Z-linked genes have a mean GC3% of 41% compared to 67% for autosomal sequences. These figures are highly significantly different (Mann-Whitney U test, P=0.0006). Our figures appear to be consistent with analysis of all 1454 chicken coding sequences described in GenBank. The mean GC3% for these is 60.4% (from www.dna.affrc.go.jp/~nakamura/CUTG.html). It is to be expected that our autosomal figure should be above this as some (unknown) percentage of the 1454 genes are Z-linked.

Only four sequences with putative open reading frames have been described on the chicken W-chromosome. With a mean GC3% of 34.7%, these have, as expected, a mean GC3 content lower than both Z-linked and autosomal genes. Whilst, however, this figure is significantly lower than the autosomal figure (Mann-Whitney U test, P=0.0009), it is not significantly different from the sequences on the Z-chromosome (Mann-Whitney U test, P=0.149), although with a total sample size of only 11, this should not be taken as a strong rejection.

We conclude that, at least in some warm blooded vertebrate species, there is a significant tendency for introns to be smaller in GC rich regions. We have failed to reject the hypothesis that in these species the recombination rate also positively covaries with GC content. We cannot hence reject the hypothesis that recombination explains some (but possibly not much) of the intron size variation, but we cannot know whether this is the result of stronger selection associated with recombination or a mutational bias. Analysis of the recombination pattern in cold blooded species will provide a further test of the proposed link between recombination and intron size.

Acknowledgements: We thanks two anonymous referees for their comments and Laurent Duret for access to unpublished data.

References

1 Duret, L., Mouchiroud, D. and Gautier, C. (1995) Statistical-analysis of vertebrate sequences reveals that long genes are scarce in GC-rich isochores. *J Mol Evol* 40, 308-317

2 Eyre-Walker, A. (1993) Recombination and mammalian genome evolution. *Proc. R. Soc. Lond. B* 252, 237-243

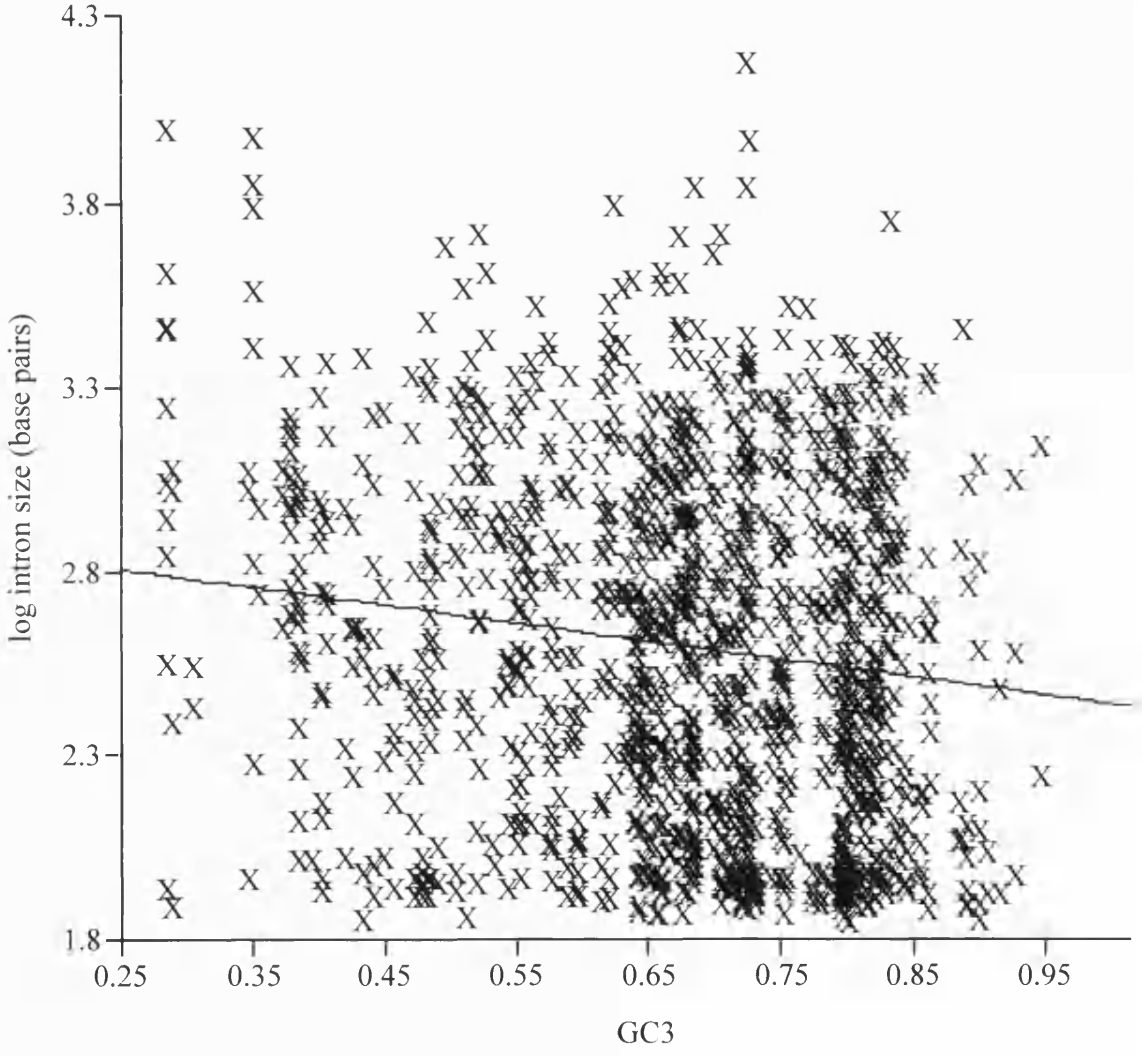
3 Nordborg, M., Charlesworth, B. and Charlesworth, D. (1996) The effect of recombination on background selection. *Genet. Res.* 67, 159-174

4 Duret, L., Mouchiroud, D. and Gouy, M. (1994) Hovergen - a database of homologous vertebrate genes. *Nucl. Acid. Res.* 22, 2360-2365

5 Bernardi, G., Hughes, S. and Mouchiroud, D. (1997) The major compositional transitions in the vertebrate genome. *J. Mol. Evol.* 44, S44-S51

6 Ogata, H., Fujibuchi, W. and Kanehisa, M. (1996) The size differences among mammalian introns are due to the accumulation of small deletions. *Febs Letters* 390, 99-103

Figure 1. Regression of the gene GC3% content against log of intron size for 1211 *Homo sapiens* introns. GC3% negatively correlates with intron size $P<0.0001$. To some extent this result is influenced by an abundance of especially small introns at especially high GC3 levels. However, restricting analysis to those with GC3% less than 70% stills reports a significant correlation albeit one of a lesser order ($P=0.013$).



Chapter 9. Does genomic anatomy make sense?

This thesis is entitled "Searching for Sense in the Library of Babel". In Chapter 1 I used the analogy of a library to illustrate how to define genomic anatomy in the terms of copy number, arrangement, and identity. I also explained that by "sense" I meant "an explanation in terms of evolutionary history", which requires deterministic evolutionary forces such as selection or mutation. Having set out the terms of the thesis, I then searched for sense in genomic anatomy from the scale of the genome down to the scale of the individual nucleotide. In this chapter I shall review my findings and discuss to what extent we know whether genomic anatomy makes sense.

Do ploidy levels make sense?

At the level of the genome the principle question of genomic anatomy is one of copy number: why do ploidy levels vary? Multiple copies of the haploid genome are unlinked (arrangement) and very similar to one another (identity). A number of ecological arguments have been put forward to explain ploidy levels, but the proposed processes are hard to evaluate (for experimental difficulties see Mable and Otto 1998). The more quantitative arguments employ evolutionary genetics. For asexual organisms haplonty is favoured by deleterious mutations, although diplonty is favoured if advantageous mutations are at least partially dominant.

For sexual organisms, the situation is more complex. Load arguments suggest that deleterious mutations favour haplonty unless truncation like selection operates. However, modifier analysis shows that deleterious mutations favour diplonty when penetrance is low, when selection is weak, when recombination is high, and when somatic mutations are important. Haplonty is favoured by inbreeding, which reduces recombination, and intraorganismal selection, which reduces the haploid mutation rate. Advantageous mutations favour diplonty when penetrance is high and when recombination is high.

Most theories concerning ploidy levels favour either haploidy or diploidy, rather than haploid-diploid life cycles. Several theories of disruptive selection have been suggested to explain haploid-diploidy, and the theory of disruptive reproductive selection appears to be supported by comparative data, but as with ecological arguments such theories are difficult to test.

Polyploidy is found in both asexual and sexual organisms. The greater prevalence of somatic, as opposed to germline, polyploidy is consistent with selection against deleterious mutations in the soma. In sexual organisms, polyploidy is not stable and seems to decay back down to diplonty. The evolutionary advantage of polyploidisation appears to derive from the increase in genetic variability which allows adaptive evolution and the acquisition of novel traits. The evolution of vertebrate development supports this explanation, since a number of tetraploidisations appear to have driven the great adaptive radiation of the vertebrates.

Haplodiploidy represents sex-specific ploidy level variation. A number of theories have attempted to explain the origins of haplodiploidy: conflict over the sex ratio, maternal genome drive, and deleterious mutations. I have shown how the latter two arguments are both affected by the level of inbreeding (Research Paper 1).

Does the evolution of genes make sense?

I have addressed the genomic anatomy of genes by considering two ways in which gene copy number can change. Copy number can increase and thus multiple gene copies can evolve. Alternatively, effective copy number can decrease, as happens in the evolution of monoallelic expression.

Increases in gene number are thought to be the main mechanism for increasing developmental complexity. The consequences of gene duplication are dependent on dosage sensitivity and the fitness effects of mutations. The C value paradox resolves to the issue of the fitness effects of replicates on individual organisms: if replicates are advantageous then genome size is adaptive, if replicates are neutral then genome size differences are due to junk DNA, and if the replicates are disadvantageous then genome size differences are due to the spread of selfish DNA.

Given the presence of multiple copies, how are they arranged and how similar are they? The process of concerted evolution increases similarity between multiple copies. Concerted evolution makes sense from a mutational viewpoint, but I have shown that concerted evolution can also be adaptive under certain conditions (Research Paper 2). These conditions are the same as those which favour haplonty over diplonty and depend on dosage sensitivity, strength of selection, and linkage.

It is not clear whether the current arrangement of genes within the genome is the result of selection, of neutral mutations and genetic drift, or a lack of mutation reflecting historical effects. There is some evidence for positive selection on gene arrangement, but the generality of such selection is unclear. I have attempted to discriminate various theories for the evolution of homologous clusters (Research Paper 3). In the absence of sufficient knowledge of the fundamental factors - mutation processes, selection processes, and the prevalence of homologous clusters - hypothesis testing is difficult.

I have examined two cases of monoallelic expression: X-linked genes and autosomal imprinted genes. The monoallelic expression of X-linked genes appears to be the consequence of a complex evolutionary history dominated by selection. There are also several selective explanations of imprinting. I have tested the conflict theory of imprinting by looking for signs of arms races in sequence comparisons (Research Paper 4), and have demonstrated that stabilising selection has affected *H19* (Research Paper 5).

Evolutionary explanations can be divided into those which invoke selection as the driving force of evolution and those which invoke mutation. The theory of adaptive mutation rates

suggests that selection and mutation explanations are not mutually exclusive since selection can adjust mutation rates. The theory suggests that both X-linked genes and imprinted genes should enjoy lower mutation rates than biallelically expressed autosomal genes as a result of their hemizygous expression. I have confirmed this prediction by sequence analysis (Research Paper 6). My analysis has also suggested the existence of a limit to selection on the mutation rate in mammals: the mutation rates of whole chromosomes or clusters of genes can be adaptively adjusted, but selection appears too weak to affect the mutation rate of a single gene.

Knockout phenotype provides a direct measure of dispensability, Q, the probability that an organism can survive and reproduce without a certain protein, and thus enables a test of Wilson *et al.*'s (1977) prediction that the rate of protein evolution depends on Q. However I have shown that the analysis of protein evolution, as assayed by K_A/K_S, reveals that knockout phenotype does not covary with evolutionary rates once the effects of tissue specificity are removed (Research Paper 7). It is unclear why this is the case, since a number of potential explanations can be envisaged.

Does the evolution of nucleotides make sense?

It often appears impossible to make evolutionary sense of individual nucleotides – why is there an A in the second intron of the human *Igf2r* gene 132 base pairs downstream of the intron-exon boundary? - although some clear examples do exist. The GT-AG rule of intron boundaries reflects a selective explanation of nucleotide identity. The presence of a TpG dinucleotide pair where the ancestral sequence is CpG suggests a methylation-induced mutational explanation.

The fact that the genetic code is degenerate means that nucleotides at synonymous sites within a protein coding sequence cannot be explained by selection on amino acid sequence. How then can we make sense of the nucleotide identity of such sites? If the mutation-selection-drift theory of synonymous codon usage is true, we cannot explain nucleotide identity at individual sites, but we can explain general patterns of nucleotide identity. Therefore the search for sense becomes an examination of the relative strengths of selection and mutation in the evolution of synonymous sites.

In *E. coli* and *Drosophila* selection at silent sites is powerful enough to bias codon usage away from the mutational expectation. But in mammals selection is unlikely to be as powerful due to smaller effective population sizes. I have described a number of studies which address the question of whether silent sites in mammals are effectively neutral (Research Papers 4, 6 and 8). I have found no evidence for selection on codon usage, but I have obtained preliminary indications that silent sites at the start of mammalian genes are subject to functional constraints, perhaps because of selection on RNA structure (Research Paper 8).

The evolutionary origins of introns are still unclear, although the IL (introns late) theory appears to be in the ascendancy. Under this hypothesis, introns have spread relatively recently as the result of selfish horizontal transmission through the genome and the weakness of selection against introns. The hypothesis that introns are not subject to purifying selection can be tested by a

comparison of intronic and synonymous substitution rates (Research Paper 9). Unfortunately such tests assume that substitution rates can be reliably estimated, but this does not appear to be the case for diverged non-coding sequences owing to problems of alignment (Research Paper 11).

Intron length appears to covary with base composition in mammals. This relationship can be explained by recourse to explanations invoking either mutation or selection, although the discrimination of the two alternatives does not yet appear to be possible (Research Paper 10).

Fundamental evolutionary factors

In many cases it is possible to make theoretical predictions of evolutionary sense, but for such models to be useful we need to understand more about the fundamental simplifying parameters of population genetics. I shall briefly review what is known about these fundamental evolutionary factors.

Mutations provide the raw material of evolution. Advantageous mutations are required for adaptive evolution, but deleterious mutations also have important evolutionary consequences (Charlesworth and Charlesworth 1998; Kondrashov 1998). According to Kimura (1983), it was H.J.Muller who first established the principle that most mutations are probably deleterious. Fisher (1930) used the analogy of adjustments to a microscope to suggest that most mutations of large effect would be deleterious whereas for mutations of very small effect the proportions of deleterious and advantageous mutations would be roughly equal. However, the process of adaptive mutagenesis in bacteria which generates mostly advantageous mutations suggests that mutations need not be blind to their potential effects (Hall 1998).

The true distribution of fitness effects of mutations is not known (Elena *et al.* 1998; Keightley 1994), and provides the cornerstone of the ongoing neutralist-selectionist debate. There are many examples of adaptive evolution at the molecular level (Gillespie 1991; Golding and Dean 1998), but such anecdotal evidence only serves to show that advantageous mutations do sometimes occur, a rather unsurprising finding given that life has managed to survive on this planet for over a billion years.

The measurement of mutation rates is difficult because they are so low, thus sequence comparison rather than direct observation is often used (as in Eyre-Walker and Keightley 1999). The experimental measurement of the selective effects of mutations is very difficult for mutations of small effect (but see Brookfield 1997a; Elena *et al.* 1998), so the use of sequence analysis (Akashi 1999) may be helpful. Mutation rates vary across the genome: on a regional basis (Casane *et al.* 1997), according to local effects (Bulmer 1986), and even dependent on methylation status (Kendrew 1994). Some of this variation may be adaptive, as in the low mutation rate of the X chromosome in mammals (McVean and Hurst 1997a). Another complication is that mutation rates may differ between the sexes. Although differences in the number of germline replications do seem to cause sex biases in mutation rates, this explanation does not explain all of the sex bias data

(Hurst and Ellergren 1998). Finally there are many different sorts of mutational processes, which will probably differ in their rates and effects.

The concepts of dominance, epistasis, and pleiotropy are all related to the effect of gene copy number on fitness, and so are of obvious importance to the evolution of genomic anatomy. The quantification of these terms is complicated by the fact that dominance is defined by phenotype rather than gene (Porteous 1996).

The relative phenotypes of heterozygotes and homozygotes are described by dominance relations, with the penetrance defined as the ratio of the effect of a heterozygous mutation to the effect of a homozygous mutation. Penetrance in *Drosophila* depends on the strength of selection (Simmons and Crow 1977). Lethals are highly recessive ($0.01<h<0.03$) while mildly deleterious mutations show greater dominance ($0.3<h<0.5$). How can this tendency to recessivity be explained?

Fisher (1928) and Haldane (1930) suggested that selection could favour modifiers of dominance which would ensure that most mutations were recessive. Wright (1934) doubted whether selection would be powerful enough to affect dominance relations, and suggested the answer to the problem might lie in the physiology of gene action. By elegant experiments Orr (1991) showed that mutations in haploids tend to recessivity just as much as mutations in diploids, hence supporting Wright over Fisher and Haldane.

Kacser and Burns (1981) agreed with Wright in considering dominance as a biochemical problem: the tendency to recessivity comes from the small flux control coefficients for single steps in a metabolic network (see Porteous 1996). It is possible, however, that selection might act on dominance via alterations to the structure of a metabolic network. Or perhaps flux control coefficients are not necessarily small, but it is small flux control coefficients which are preferentially measured (Mazat *et al.* 1996). Although the biochemical explanation of recessivity appears to work for deleterious mutations affecting enzymatic pathways, it is not clear whether the results can be extrapolated to other situations (Bourguet and Raymond 1998).

The phenomenon of overdominance or heterozygote advantage provides a selective explanation for high levels of polymorphism. The neutral theory proposes an alternative mutational explanation and in *The Neutral Theory of Molecular Evolution* (1983) Kimura marshals an impressive body of evidence against overdominance and concludes "Evidence against the overdominance hypothesis of protein polymorphism has accumulated so much now that only blind faith can maintain it". The phenomenon of heterosis can be attributed to either overdominance or inbreeding depression (Parsons and Bodmer 1961), although discriminating the alternative hypotheses is difficult (David 1997). The nature of the correlation between recombination and neutral variation suggests that dominance rather than overdominance is the major genomic mode in mice and humans (Deng *et al.* 1998).

Epistasis describes how genes work together to affect organismal fitness, and has profound implications for evolutionary biology (e.g. Kondrashov 1988). Elena and Lenski (1997) have

shown that in *E. coli* epistasis is widespread with interactions between half of all pairs of mutations, results which are consistent with evidence in other organisms (Otto 1997). As with dominance, the nature of epistasis can be considered to derive from metabolic control (Kacser and Burns 1981; Szathmáry 1993).

Pleiotropy measures the extent to which mutations in one gene can have multiple effects, and is dependent on the nature of metabolic pathways (Kacser and Burns 1981). Antagonistic pleiotropy is considered important in evolutionary ecology, although it seems a weak force for the maintenance of genetic polymorphisms (Hedrick 1999).

Recombination plays a central role in theoretical evolutionary genetics, but little is known of the patterns of recombination across the eukaryotic genome (Comeron *et al.* 1999; Nachman and Churchill 1996). The evolution of breeding systems and hence levels of inbreeding has been well studied, especially with regard to the effects of deleterious mutations (Charlesworth and Charlesworth 1998). It is possible to measure absolute population sizes with reasonable accuracy, but the measurement of the theoretically more important effective population size is considerably more difficult, being dependent on additional factors such as sex ratio bias, variable offspring production, temporal variation in absolute population size, overlapping generations and population sub-structure (Kimura 1983).

Conclusions

If we want to know whether genomic anatomy makes sense, we should test simple models which assume that the appropriate mutations can occur (in other words, assume no strong constraints). For example, in my model of the evolution of haplodiploidy I assumed a perfectly functional and costless haplodiploidy modifier (see Research Paper 1). Using such an approach the results are dependent on the underlying parameters of population genetics. Many of the conclusions reached in this thesis are subject to assumptions concerning the fundamental factors reviewed above: mutation rates and fitness distributions, dominance, epistasis, pleiotropy, recombination, breeding systems and population size. These factors represent simplifications of the real world, but they do represent ubiquitous evolutionary forces rather than the specific biological details invoked by constraints.

So in order to find out whether genomic anatomy makes sense, and to discover whether mutation or selection is responsible, we need to find out more about the real distributions of the parameters we use in our necessarily simplified models of the evolution of the genome. I conclude with a quotation from Kimura (1983): "Among students of evolutionary biology, there has been a strong tendency to claim that these population genetical parameters will never be known accurately and therefore theories which contain such parameters are of little use. I take the opposite view; these parameters have to be investigated and measured if we really want to understand the mechanism of evolution at the molecular level".

Chapter 10. Hypothesis testing in molecular evolution

The general approach of hypothesis testing in molecular evolution relies on statistical inference. For hypothesis testing to be valid, a number of requirements must be satisfied. Firstly, the sequences to be analysed must constitute independent evolutionary experiments. Secondly, the sequence data must be analysed correctly to yield unbiased quantitative data. Finally, the *a priori* predictions of the hypothesis must be tested using the appropriate statistical methods (under probabilistic, rather than likelihood based, inference).

My interest in the importance of methodology in evolutionary inference was first prompted when I attempted to test silent site neutrality by comparing rates of evolution at synonymous and intronic sites. There seemed to be no reliable way to choose between different intronic alignments produced by different settings of the same alignment program, despite my efforts to develop suitable methods (see Research Paper 9). Could such problems of methodology apply to other aspects of sequence analysis? I investigated this problem by examining the sensitivity of the results of a single paper (Hughes and Yeager 1997) to alterations in methodology (see Research Paper 11). In this chapter I shall consider some of the many ways in which methodology can affect evolutionary inference in molecular evolution.

Independence

Statistical methods assume the independence of data and so non-independence of sequences can lead to biased P value estimates. There are two principle causes of non-independence: ascertainment biases and shared evolutionary history.

The presence of ascertainment bias can be demonstrated by significant correlations between details of dataset assembly, such as sample size, and sequence analysis details of the dataset, such as mean intron size. The comparative method (Harvey and Pagel 1991) takes account of shared evolutionary history when analysing species specific characters. The problem of shared evolutionary history also applies to molecular evolution studies which involve large numbers of pairwise comparisons, especially if all pairwise comparisons are assumed to be independent (as in Wolfe and Sharp 1993). This assumption of independence may be flawed for a number of reasons.

The need to compare orthologs rather than paralogs is ensured by checking phylogenetic trees in homology databases (Duret *et al.* 1994; Makalowski *et al.* 1996), although such methods assume correct tree reconstruction and complete sequence data. Another problem affecting gene families is concerted evolution, which is demonstrated by greater divergence for between species comparisons than for within species comparisons (as for the β globin genes in Research Paper 11). The existence of multimeric genes like *Igf2r* (see Research Paper 4) can lead to concerted evolution within a gene.

The possibility of shared evolutionary history also applies within genes. Should the exons and introns of a single gene be considered as separate data, or should they be combined into a single datum? It is sometimes necessary to separate exons and introns (as in the analysis of *H19* in Research Paper 5) but otherwise the introns and exons of a single gene which have been subjected to very similar evolutionary forces should probably be combined.

The problem of independence leads on to issues of weighting. Should data points be weighted by gene length? Longer sequences may give more reliably estimated data, but confounding factors might correlate with gene length and shorter genes may be more important than long genes.

Sequence alignment

The need for sequence alignment arises when one wishes to compare homologous sites in two sequences: one has to first account for the insertion and deletion events which have taken place since the two sequences diverged (Li and Graur 1991). Most alignment algorithms take a parsimony approach: either the similarity between the two sequences is maximised or the distance between the two sequences is minimised by dynamic programming (Li 1997). Which alignment is best is determined by the parameter settings. Typically parameters determine the scores for matches, mismatches, indel creation, and indel extension. The problem of alignment lies in the fact that we have no reliable way of determining which alignment parameters are best (Altschul 1997)(see Research Paper 9).

The choice of parameters holds clear consequences for the estimation of substitution rates, since the parameters determine the relative frequencies of matches and mismatches. Two popular alignment programs yielded significantly different estimates of intronic substitution rates when the default parameter settings were used (Research Paper 11). In contrast, the exonic alignments were nearly identical. Exonic alignments are easier to perform because exons evolve more slowly than non-coding DNA, and because protein and DNA alignments can be combined.

Sequence alignment also affects the accuracy of tree reconstruction (Goldman 1998). One possible solution to these problems is a combined alignment, rate estimation and tree reconstruction approach based on maximum likelihood, but the computational demands are likely to be formidable.

Substitution rates

There exists a wide variety of methods for estimating substitution rates from DNA sequence alignments (see Chapter 4 in Li 1997). These methods differ in the complexity of their models of nucleotide substitution. There are two main classes of method: multiple substitution correction applies to both protein coding and noncoding DNA sequences, while the calculation of synonymous and nonsynonymous substitution rates applies to protein coding sequences only.

In general, the more complicated the model of nucleotide substitution, the greater the number of free parameters, the better the estimates of substitution rates. But Zharkikh's (1994) simulation study pointed out that all methods are unreliable when divergences are high (K>1) and that more complicated methods are often inapplicable for short sequences. More information is required to estimate more parameters, and so the estimates of better methods are likely to be more variable.

Multiple substitution correction methods differ in how they account for variation in base composition. If equal base frequencies are assumed substitution rates will be underestimated when composition is biased (Pesole *et al.* 1995). For example, the demonstration of a correlation between GC and K_S (Wolfe *et al.* 1989) was thought to be an artefact of such biased substitution rate estimation (Bernardi *et al.* 1997). However, the finding of a correlation between GC and K_S using the maximum likelihood program PAML complicates the issue (see Research Paper 8).

Some of the older methods of synonymous and nonsynonymous rate estimation (such as that of Li *et al.* 1985a) fail to account for different rates of transition and transversion, which leads to an overestimation of K_S by about 30% (Li 1993). Clearly, the use of different methods can lead to very different results when synonymous substitution rates are compared to other substitution rates, as shown by my comparison of K_S and K_I using alternative methodologies (see Research Paper 11).

Simple methods assume that all sites within a gene evolve at the same rate, apart from sites of different degeneracy. However, there is ample evidence that rates vary within genes (Yang 1996), for example there is deterministic variation in both synonymous and nonsynonymous rates across the *Igf2r* gene (see Research Paper 6). Current methods use the gamma distribution (or a discrete approximation) to model rate variation. The problem of choosing the shape of the gamma distribution is circumvented by maximum likelihood techniques, in which the likelihood of observing the data given the model parameters is maximised. Although the maximum likelihood approach can be used to generate synonymous and nonsynonymous rates, at present it does not appear possible to incorporate rate variation into such estimates (Z.Yang, pers. comm.). The gamma distribution does not seem suitable for a codon based model, and there is also the problem that specifying a codon as fast evolving automatically generates an intragenic KA-KS correlation which may not be justified (Z.Yang, pers. comm.).

Statistical methods

The concomitant variation of observations is tested using the statistical techniques of correlation and regression, although care must be used in the interpretation of parametric statistics whose assumption of a particular underlying distribution may not be justified. For example, in Research Paper 11 I describe attempts to quantify the strength of the relationship between the base composition of mouse and rat genes at different codon positions. Although it was clear on observing plots of the data that the relationship was much tighter for GC at the first and second

codon positions than for GC at the third codon positions, the parametric linear regression P values appeared to show the converse. However when the residuals of the linear regression were compared by the nonparametric Mann-Whitney U test it was shown that the residuals were higher for the third codon position, which confirmed the visual analysis. The discrepancy seems to be due to the fact that linear regression is sensitive to outliers. A couple of outliers in the GC3 data appeared to generate an artificially strong regression slope.

The confounding effects of outliers was shown even more strikingly in the same study by measurements of the K_A-K_S correlation. Using parametric linear regression Hughes and Yeager (1997) found no significant K_A-K_S correlation, in contrast to a number of previous studies (such as Wolfe and Sharp 1993). I found that the removal of one outlier out of 41 genes caused a 80 fold drop in the P value measured by linear regression: the correlation became highly significant. The relative insensitivity of nonparametric methods to outliers was confirmed by the use of rank correlation which revealed a significant correlation both in the presence and the absence of the outlier.

Research Paper 11. Sensitivity of patterns of molecular evolution to alterations in methodology: a critique of Hughes and Yeager

Nick Smith and Laurence Hurst (1998)
J. Mol. Evol. **47** 493-500.

J Mol Evol (1998) 47:493–500

JOURNAL OF MOLECULAR EVOLUTION

Letter to the Editor

Sensitivity of Patterns of Molecular Evolution to Alterations in Methodology: A Critique of Hughes and Yeager

Nick G.C. Smith, Laurence D. Hurst

Centre for Mathematical Biology, School of Biology and Biochemistry, University of Bath, Claverton Down, Bath BA2 7AY, England

Received: 2 February 1998 / Accepted: 23 March 1998

Abstract. Employing a set of 43 othologous mouse and rat genes, Hughes and Yeager (*J. Mol. Evol.* 45:125–130, 1997) reported (1) no correlation between synonymous and nonsynonymous rates of nucleotide substitution, (2) a positive correlation between intronic GC contents (GC_i) and intronic substitution rates (K_i), (3) that the average K_i value was very similar to the average K_s value, and (4) that the compositional correlation between the rat and the mouse genes is stronger at the third codon position (GC_3) than at the first and second codon positions (GC_{12}). We have examined the robustness of these results to alterations in substitution rate estimation protocol, alignment protocol, and statistical procedure. We find that a significant correlation between K_a and K_s is observed either if a rank correlation statistic is used instead of regression analysis, if one outlier is excluded from the analysis, or if a regression weighted by gene size is employed. The correlation between K_i and GC_i we find to be sensitive to changes in alignment protocol and disappears on the use of weighted means. The finding that K_s and K_i are approximately the same is dependent on the method for estimating K_s values. Finally, the variance around the regression line of rat GC_3 versus mouse GC_3 we find to be significantly higher than that in GC_{12}. The source of the discrepancy between this and Hughes and Yeager's result is unclear. The variance around the line for GC_4 is higher still, as might be expected. Using a methodology that may be considered preferable to that of Hughes and Yeager, we find that all four of their results are contradicted. More importantly this analysis reinforces the need for caution in assembling and analyzing data sets, as the degree of sensitivity to what many might consider minor methodological alterations is unexpected.

Key words: Substitution rate estimation — Intron evolution — Synonymous and nonsynonymous correlation — GC composition

Correspondence to: N.G.C. Smith; *e-mail:* n.smith@bath.ac.uk

Introduction

Hughes and Yeager (1997) (HY97) have assembled a set of 42 othologous mouse and rat genes for which both intronic and exonic sequence data are available. As their method for assigning orthology is especially stringent, this data set provides a potentially valuable tool for the analysis of patterns of molecular evolution of introns and exons, as any resulting patterns can, possibly for the first time, be guaranteed not to be the result of contamination of the data set with paralogous genes.

Among other results, the pair reports four important findings, namely, (1) no correlation between synonymous and nonsynonymous rates of nucleotide substitution (K_s and K_a, respectively), (2) a positive correlation between intronic GC contents (GC_i) and intronic substitution rates (K_i), (3) that the average K_i value was very

similar to the average K_s value, and (4) that the correlation of the third codon position composition (GC_3) between the rat and the mouse genes is significantly stronger than the correlation of the first and second position composition (GC_{12}).

The first result is surprising given that several previous studies found a significant correlation between K_a and K_s (for references see HY97). HY97 suggested that previous results may be misleading, either because of the presence of paralogues or because of biased substitution rate estimation procedures. Were the latter true, then different substitution rate estimation protocols should find a significant correlation between K_a and K_s in the HY97 data set. We examine this issue here.

In addition, there are at least two other components of the methodology that might affect the results, namely, alignment method and statistical analysis. We hence examined the robustness of Hughes and Yeager's results not only to alterations in the substitution rate estimation protocol, but also to changes in the alignment protocol and statistical procedure. Perhaps surprisingly, we find that three of HY97's results are sensitive to methodological alteration, which illustrates the need for caution when choosing methodology.

Methods

We used the GCG (Genetics Computer Group 1994) and EGCG (Rice 19XX) sequence analysis packages to prepare our alignments. FETCH was used to obtain the DNA sequences with the accession numbers for the 42 genes provided by HY97. Due to a minor typographical error, one gene in HY97 has the wrong accession number. We confirmed with the authors that we were using the correct one.

Introns were manually extracted using SEQED. GENETRANS was used to extract the protein coding sequences, and we checked all sequences for appropriate start and stop codons. This method is slightly more stringent than that of HY97 in that we use only independent data on exon–intron boundaries to determine the coding sequence. This led to the exclusion of the kallikrein binding protein gene from our data set since no such information was available for the rat gene. This otherwise is the only difference in the source data used between our and HY97's analysis.

Both intronic and exonic alignments were prepared using PILEUP and CLUSTALW. Both programs' default settings were used for one set of alignments, and a further set of alignments was obtained for both programs by penalizing terminal gaps (which are not penalized by default). Exonic alignments were corrected for possible frameshifts by GAPFRAME. This automated procedure was used rather than "correction by eye" to avoid possible bias. Alignments of all the introns were produced using the default settings of the GAP program and were examined so that all introns showing large indels could be discarded from the analysis. A similar approach was used by HY97 since alignments containing large indels are unreliable. Thus four alignments were obtained for 41 exonic and 120 intronic sequences.

There are many distance estimation protocols for both protein-coding and noncoding sequences (see Li 1997, Chap. 4). Methods for both protein-coding and noncoding sequences differ in terms of accounting for variable sequence composition and, thus, produce different distance estimates when sequence composition is biased (Pesole et al. 1995). Some of the methods for protein-coding sequences fail to take into account the transition/transversion bias and the fact that most transitional mutations at twofold degenerate sites are synonymous, which leads to a systematic overestimate of K_s (Li 1993). We used 14 methods for analyzing protein-coding sequences and 9 methods for noncoding sequences. These 9, and 13 of the 14 methods, were as implemented in the Molecular Evolutionary Analysis package obtained from Moriyama (Moriyama and Powell 1997) (see Table 1). The same package was used to assess GC content. In addition, we used Ina's (1995) package as the fourteenth method for analyzing protein-coding sequences.

All gene-specific measures (K_a, K_s, K_i, and GC_i) of HY97 are unweighted means, i.e., they calculated means by using each intron and exon as a single data point, making no allowance for relative sizes. In contrast, we use weighted means for all gene parameters, i.e., exons and introns are scored according to their size. The K_s and K_a values we obtained were automatically weighted means for exonic size, since all the protein-coding exons were combined during sequence extraction. The K_i values were initially obtained for individual introns. The individual introns of each gene were weighted by length (number of bases compared) and a weighted mean K_i was calculated for each gene.

The Correlation Between K_a and K_s

Many previous analyses have found a correlation between K_a and K_s and many reasons have been given to explain such a result (see Li 1997). It was thus surprising to find that an analysis with such strong controls on paralogy should not find the effect. If linear regression analysis is performed on the corrected K_a and K_s values from HY97 and if the kallikrein-binding protein gene is excluded from the analysis, no significant correlation is found ($P = 0.186$). We have examined the robustness of this result to changes in alignment protocol, distance estimation, and statistical method.

The Effect of Alignment Method. The four alternative alignment protocols produced very similar sets of exonic alignments, which yielded correspondingly similar distance estimates. Using Li's (1993) method with Tamura and Nei's (1993) (TN93) correlation for multiple substitutions, all four alignment protocols showed no significant K_a/K_s correlation with weighted regression analysis (P values ranged from 0.132 to 0.142). Thus HY97's result of no K_a/K_s correlation seems insensitive to changes in alignment protocol.

The Effect of Distance Method. Using the default PILEUP exonic alignments, the K_a and K_s values of 14 distance estimation methods were tested using linear regression analysis. Only 4 of the 14 methods supported a K_a/K_s correlation at the 5% level, and the mean regression P value was 0.08 (see Table 1). So distance estimation methodology can have some effect, but HY97's result is in agreement with the majority of distance estimation methods.

The Effect of Statistical Method. The same set of values (default PILEUP alignment and 14 distance estimation methods) were analyzed using the nonparametric statistical method of rank correlation. Under this statis-

Table 1. *P* values for the K_a/K_s correlation, using different distance estimation methods, different statistical tests, and different sets of data[a]

Distance method	Statistical method and data set					
	P, all-unwt	P, all-wt	NP, all-unwt	P, sub-unwt	P, sub-wt	NP, sub-unwt
NG86 Unc	0.052	0.002	<0.02	<0.001	<0.001	0.001
NG86 JC69	0.056	0.001	<0.02	<0.001	<0.001	0.001
NG86 K80	0.064	0.002	0.01	0.001	<0.001	0.001
NG86 TN84	0.096	0.004	>0.05	0.001	<0.001	0.01
NG86 T92	0.086	0.003	<0.05	0.001	<0.001	0.01
NG86 TN93	0.1	0.004	<0.05	0.002	<0.001	0.01
Com95 K80	0.002	<0.001	<0.001	<0.001	<0.001	<0.001
Li93 K80	0.099	0.002	<0.05	<0.001	<0.001	0.005
LWL85 K80	0.048	0.001	0.01	<0.001	<0.001	0.001
Com95 T92	0.087	0.002	0.02	<0.001	<0.001	0.005
Li93 T92	0.115	0.002	<0.05	0.001	<0.001	<0.01
Com95 TN93	0.11	0.001	0.02	0.001	<0.001	0.005
Li93 TN93	0.131	0.003	<0.05	0.001	0.001	0.01
Ina95 K80	0.031	<0.001	<0.005	<0.001	<0.001	<0.001
Mean	0.08	0.002	~0.004	~0.001	~0.001	~0.004

[a] In the "distance method" column, the protein-coding sequence distance method is given first, followed by the multiple substitution correction method. NG86, Nei and Gojobori (1986); Com95, Comeron (1995); Li93, Li (1993) and Pamilo and Bianchi (1993); LWL85, Li et al. (1985); Ina95, Ina (1995); Unc, no multiple substitution correction was used; JC69, Jukes and Cantor (1969); K80, Kimura (1980); TN84, Tajima and Nei (1984); T92, Tamura (1992); TN93, Tamura and Nei (1993). Under "statistical method and data set," P stands for the parametric method of linear regression, and NP for the nonparametric method of rank correlation. "sub" means the exclusion of *Interleukin 3* from the analysis, whereas "all" means it was included. "wt" means the regression was weighted by gene size, whereas "unwt" means that although each gene's values are weighted by exon size, in the regression gene size was not a weighting parameter. We did not perform nonparametric analysis with weighting by gene size. Hence for each of the 14 methods we did four parametric and two nonparametric analyses.

tical analysis 13 of the 14 distance estimation methods supported a K_a/K_s correlation at the 5% level, and the mean correlation coefficient was 0.38 ($P < 0.025$) (Table 1).

One difference between the parametric regression analysis and the nonparametric rank correlation is the former's greater sensitivity to outliers. In the plot of K_a against K_s for all the genes an extreme outlier is *Interleukin 3,* which has a K_a/K_s ratio significantly greater than 1. If just this one gene is excluded from the data set and the regression analysis is repeated, then all 14 distance estimation methods show a highly significant K_a/K_s correlation, with a mean regression *P* value of 0.001 (Table 1). Thus the removal of just 1 gene of 41 leads to an 80-fold reduction in the regression *P* value, which shows that the failure of HY97 to find a significant K_a/K_s correlation was the unfortunate result of one extreme outlier and a statistical method sensitive to outliers. Nonparametric analysis of the dataset with *Interleukin 3* removed was also highly significant for all 14 methods (Table 1).

Our results on the sensitivity of HY97's K_a/K_s findings are in agreement with those of Makalowski and Boguski (1998). Using a large data set of 470 mouse and rat genes of confirmed orthology, and using estimation method of Ina (1995), they confirmed previous results of a significant positive K_a/K_s correlation ($P < 0.0005$). In addition, they, like us, identified a number of mistakes in the distance estimates of the original HY97 paper (an erratum is in press, and we have used the corrected data) and point out that *Interleukin 3* is an extreme outlier to the data set.

The Correlation Between K_i and GC_i

There has been a lengthy debate as to whether GC content and substitution rate are in any way related (see Bernardi et al. 1997). The debate is of importance, as it is central to determining whether isochores might be the result of mutational biases. While the best data sets suggest no correlation between GC content and silent substitution rate (Bernardi et al. 1997), in HY97 linear regression of GC_i and K_i suggested that such a covariance does exist ($P = 0.038$). We examined whether different methodologies would lead to nonsignificant linear regression *P* values.

The method used in HY97 has two obvious potential sources of error. First, the intron distance estimates were obtained using Jukes and Cantor's (1969) (JC69) multiple substitution correction method, which assumes equal use of all four nucleotides. Second, the GC_i values given by HY97 for each gene are the unweighted means of each gene's introns. Finally, HY97 use only one alignment method and it is informative to ask whether their choice of protocol might have affected their results.

We have hence examined the effect of (1) employing different alignments (default PILEUP and default CLUSTALW); (2) using a better multiple substitution correction, namely, TN93's method (this uses all four

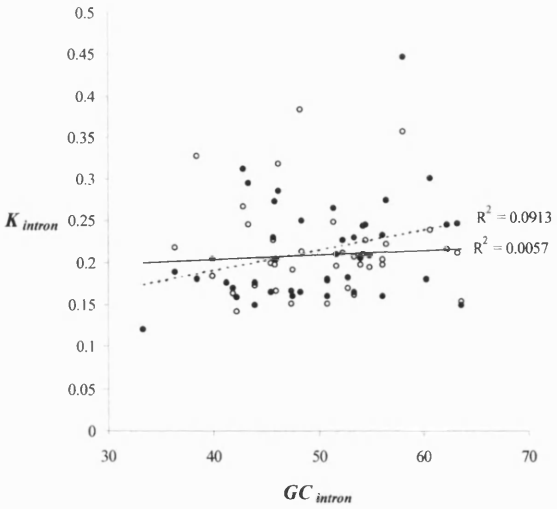


Fig. 1. Regression trend line plots for GC_i (intronic GC content) versus K_i (intronic substitution rate) for two alternative alignments protocols. Note that while PILEUP yields a linear regression line with a nearly significant positive slope, CLUSTALW gives an almost-flat linear regression plot. *Filled circles* and *dashed line* are for PILEUP, and *open circles* and *solid line* for CLUSTALW.

nucleotide frequencies, as well as a transversion substitution rate and two transition substitution rates); and (3) calculated the GC_i values for each gene as weighted means (weighted by number of bases compared), which seems more appropriate than unweighted means.

The Effect of Alignment Methods. When weighted K_i means were calculated and the TN93 multiple substitution correction method was used; PILEUP gave a P value of 0.055, while CLUSTALW gave a P value of 0.639 (see Fig. 1). The HY97 result of a significant correlation between GC_i and K_i is thus highly sensitive to changes in alignment protocol.

The Effect of Distance Method. The P value did not appear to be very sensitive to the multiple substitution correction method. For example, that of JC69 gave a P value of 0.059 with PILEUP, as opposed to 0.055, with that of TN93.

The Effect of Using Weighted Means. HY97's use of unweighted means appears to also have had a major effect on their findings. When unweighted K_i means were used, CLUSTALW gave a P value of 0.083 (as opposed to 0.639 with weighted K_i means). Therefore the HY97 result of a significant correlation between GC_i and K_i possibly should be disregarded owing to their use of unweighted means.

Is K_s Equal to K_i?

The relative size of K_s and K_i is important in determining whether silent sites might be evolving neutrally. HY97 reported that the two were more or less the same, supporting neutrality for one if neutrality for the other can be assumed or demonstrated.

HY97 estimated K_s and K_i using Nei and Gojobori's (1986) (NG86) method with JC69's correction for multiple hits. The t test indicated no significant difference between HY97's K_s and HY97's K_i values ($P > 0.75$). However, NG86's method systematically underestimates the number of synonymous sites, and hence overestimates K_s, because it fails to account for the transition/transversion bias. We find that K_s estimates using NG86's method with TN93's correction for multiple hits are, on average, over 30% higher than K_s estimates obtained using Li's (1993) (Li93) method with TN93's correction for multiple hits. The same bias will not affect intronic distance estimates since the bias appears only when protein-coding sites are divided into synonymous and nonsynonymous sites. It is, then, possible that HY97 failed to detect a difference between K_s and K_i because they systematically overestimated K_s.

We have used two alternative statistical methods to test for differences between sets of K_s and K_i data obtained using different methodologies. The parametric t test examined whether the mean difference between K_s and K_i was significantly different from zero, and the nonparametric test used the binomial distribution to see if the number of genes for which K_s exceeded K_i was significantly different from the null expectation of half the genes.

The Effect of Distance Method. With alignment and multiple correction methods constant (PILEUP default and TN93, respectively), K_s was estimated using the methods of both NG86 and Li93. With NG86 both the binomial and the t tests showed no significant difference between K_s and K_i ($P = 0.53$ and $P = 0.25$, respectively). Yet with Li93 both the binomial and the t tests showed a significant difference ($P = 0.0015$ and $P = 0.0021$, respectively), with K_i greater than K_s. Thus HY97's result of no significant difference between K_s and K_i appears to be a consequence of their distance estimation methodology.

The Effect of Alignment Method. While alternative alignment protocols give near-identical results for protein-coding sequences, a significant difference between the default intronic alignments of CLUSTALW and PILEUP was observed. For both sets of alignments, distances were estimated by the method of TN93. The t test failed to show a significant difference between the two sets of K_i values ($P = 0.47$), but the binomial test showed a significant difference, with PILEUP giving the larger K_i values for 28 of the 41 genes ($P = 0.028$).

Despite the difference between the alignment protocols, both CLUSTALW and PILEUP produce the same answer (and a different answer from HY97's) to the

Table 2a. Adjusted values for r^2 from the regression of GC content in the mouse–rat comparison for measures of five GC parameters of genes[a]

	Regression	Regression of ranks
GC_1	0.933	0.922
GC_2	0.968	0.905
GC_{12}	0.908	0.852
GC_3	0.931	0.927
GC_4	0.896	0.843

[a] Data for both a regression on the GC content and a regression on the ranking of GC content are shown. For analysis of differences in fit to the regression line see Table 2b.

Table 2b. Analysis of the residuals of the regression analysis[a]

	P value			
	GC_2	GC_{12}	GC_3	GC_4
GC_1	0.68	0.6	0.09	0.0014***
GC_2	—	0.67	0.057	0.0002***
GC_{12}		—	0.03*	0.0002***
GC_3			—	0.12

[a] For each regression analysis (e.g., mouse GC_1 versus rat GC_1), the residual values for each data point were squared. A Mann–Whitney *U* test was then performed to compare the distributions of the squared residuals so as to ask whether the distributions were the same. The null expectation if the fit to the line is no different is that the distribution of residuals should be no different. We find that GC_4 is consistently more variable around the regression line than GC_1, GC_2, and GC_{12}. GC_3 is significantly more variable than GC_{12}.

question of whether K_s is the same as K_i when Li93 rather than NG86 is used. Using the *t* test and the distance estimation methods of TN93 and Li93, both PILEUP and CLUSTALW reject the null hypothesis of K_s equal to K_i ($P = 0.0021$ and $P = 0.0047$, respectively). Alternatively, if NG86 is used instead of Li93 (use of *t* test and TN93 held constant), then both PILEUP and CLUSTALW agree with HY97's result of a failure to reject the null hypothesis of K_s equal to K_i ($P = 0.25$ and $P = 0.29$, respectively). Thus the relationship between K_s and K_i is dependent on the distance methodology and not on the alignment protocol, although the latter does introduce detectable systematic differences.

The Correlation Between Mouse and Rat Values of GC_3 and GC_{12}

It might be expected not only that evolution at silent sites should be less constrained than evolution at the first and second sites, but also that, as a consequence, the GC content at third sites might show more variation between species (and hence a weaker correlation) than that at the first and second sites. However, HY97 report the opposite result, namely, that the correlation of GC_3 between mouse and rat is significantly stronger than that of GC_{12}.

Given that this result is not obviously that one might expect, we decided to reinvestigate this issue. In their original report, as we discovered, HY97 had wrongly excised some exons, so inducing frame shift mutations in the sequences that they analyzed. This will give misleading values for GC content at each site. Further, they used gene values unweighted by the relative sizes of exons. Could either of these have contributed to their finding? HY97 also restricted their analysis to GC_3, while the null expectation of unconstrained evolution at third sites applies more strictly to fourfold degenerate sites. We hence analyze GC_4 as well. Further, GC_{12} is a composite measure of what is happening at the first two sites. We also look at each site independently.

We find that the value of the correlation coefficient for the comparison of GC_3 in mouse and rat is, as claimed, higher than for GC_{12} (Table 2a). In contrast, the correlation of GC_4 (composition at fourfold degenerate sites) is less than the correlation of GC_{12} (Table 2a). More importantly, however, when we ask about the variation around the regression line, we find a result opposite to that reported by HY97. Namely, we find that variation around the regression line is higher for GC_3 than for GC_{12} ($P = 0.05$) (Table 2b, Fig. 2). The variance around the regression line in GC_4 is highly significantly greater than that at the first two sites (see Table 2b). HY97 do not report what statistic they used to claim that the correlation in GC_3 was significantly greater than that in GC_{12}, so we do not know whether we have replicated their procedure. However, we can confirm that the variance around the regression line is as would be expected, i.e., higher at silent sites (see Fig. 2).

Conclusion

We have shown that all of HY97's results given above are sensitive to changes in methodology. To summarize, we find the following. (1) A significant correlation between K_a and K_s is observed either if a rank correlation statistic is used instead of regression analysis or if the extreme outlier *Interleukin 3* is excluded from the analysis. (2) The proposed correlation between K_i and GC_i is sensitive to changes in alignment protocol and disappears on the use of weighted means. (3) Alignment protocol has a significant effect on K_i estimates, but the finding that K_i and K_s were roughly the same is very much more sensitive to the method for estimating K_s. (4) Variation around the regression line of mouse GC_3 against rat GC_3 is greater than in GC_{12}.

Additionally we identify a few weaknesses in the data set itself. Not only has one of the genes not had its intronic structure defined; but also it is not especially large and also contains many incomplete genes (while the complete sequences are available). Further, the set

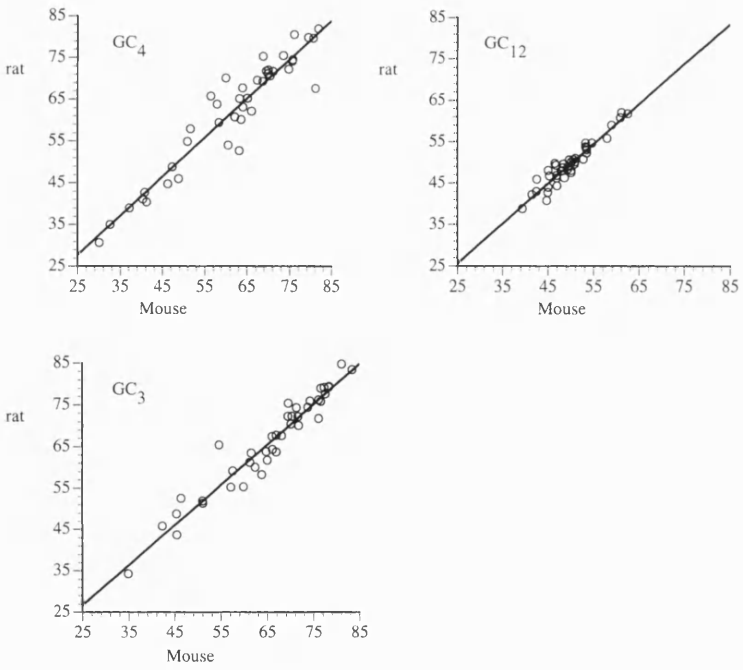


Fig. 2. The regression of GC_3, GC_{12}, and GC_4 in the mouse–rat comparison. Note the higher variance away from the regression line for GC_3 and GC_4 compared to GC_{12}.

contains both β major- and β minor-globin and treats them as though their evolution were independent. However, the distances between β major and β minor in both species are less than the distances for both genes between species, indicating concerted evolution and nonindependence (e.g., K_s mouse minor–major = 0.054, K_s rat minor–major = 0.083, K_s minor rat–mouse = 0.194, K_s major rat–mouse = 0.171; distances obtained using DIVERGE in GCG).

Our intention is not to insist that the results of HY97 should be ignored. They should, however, be treated with appropriate caution. More generally, however, we wish to bring attention to the fact that even from a data set with such tight controls on the exclusion of paralogy, a wide range of results can be obtained by application of a particular combination of methodologies. The relevance of this finding should not, however, be overstated, as some of the problems are specific to the data set (e.g., an outlier in the K_a/K_s distribution), rather than a problem with the methodology. Further, the issue is not whether we can find various results but, rather, whether some methods provide more reliable results. What, then, are the preferable methodologies, and what results are obtained if the "best" methodology is used? Here we make a few tentative recommendations.

(1) *Multiple substitution correction:* It is probably advisable to use a multiple correction method such as TN93, which takes account of variation in sequence composition, despite such methods giving large standard errors.

(2) *Protein-coding sequence analysis:* One should use an unbiased estimator of synonymous and nonsynonymous sites such as that of Li93 or latter developments thereof.

(3) *Statistical methods:* Since parametric statistical methods are more sensitive to outliers than nonparametric methods, reanalysis of subsets of the data might help to avoid methodological artifacts when using parametric statistics.

(4) *Alignment methods:* Unfortunately there is no easy way of deciding between alternative alignments protocols, since there is no theory for the selection of biologically valid alignment parameters (Altschul 1997). Although the exonic alignments of the different protocols were very similar, the different sets of intronic alignments gave significantly different results. This result casts doubt on the general method of comparing distance estimates for noncoding sequences with other distance estimates.

(5) *Calculating means:* The case in favor of gene values weighted by the relative sizes of exons and introns seems clear. However, in the above analyses we do not weight by gene size when comparing between genes. Not weighting for gene size is defensible if one argues that important parameters covary with gene size. For example, longer introns and longer genes exist in AT-rich isochores (Duret et al. 1995). When comparing between genes, then, weighting for gene size may inadvertently introduce systematic biases in that one is, by proxy, weighting for AT bias. This is not desirable when, for example, analyzing correlates of GC content or anything else that covaries with GC content. However, the same defense for weighting for exon size can be given for weighting for gene size, i.e., it is not desirable to inflate the relevance of data points that result from small samples.

On this issue, then, we remain largely agnostic. However, we do wish to note that the use of full weightings

Table 3. The set of results obtained using our designated "best" methodology[a,b]

Gene	K_a	K_s	K_i	GC_i
α–Actin	0	0.10696	0.2167	62.2
α–β$_2$–Crystallin	0.00826	0.1608	0.1994	45.4
α–Lactalbumin	0.16809	0.31324	0.2276	54.5
α–Lactid glycoprotein	0.08445	0.31905	0.2277	45.6
Anti-müllerian hormone	0.05829	0.21805	0.1814	60.2
Apolipoprotein a-i	0.1809	0.27599	0.1624	53.4
Apolipoprotein a-iv	0.09965	0.20892	0.196	51.6
Atrial natriuretic factor	0.03851	0.15021	0.1805	50.8
β Minor-globin	0.07433	0.20574	0.3274	38.5
β Major-globin	0.04755	0.17306	0.1839	39.9
C-ras-h-1	0.00697	0.08095	0.2044	56
Creatine kinase b	0.00526	0.21853	0.2123	63.2
Cytochrome c	0	0.08433	0.1665	45.9
Cytochrome P-1-450	0.04266	0.16839	0.3845	48.2
Gadd45 protein	0.01797	0.25509	0.1518	50.8
γ–C-Crystalline	0.02256	0.15991	0.2179	36.4
Gonadotropin releasing hormone	0.04949	0.13566	0.1424	42.2
Immunoglobulin light chain c-k	0.11148	0.16542	0.176	41.2
Immunoglobulin light chain v-l-1	0.11225	0.1309	0.1984	45.8
Insulin ii	0.02701	0.16001	0.1699	52.8
Interleukin 6	0.07185	0.13814	0.1646	41.9
Je	0.09832	0.09906	0.1727	44
Lymphotoxin	0.02382	0.23399	0.3578	58
Mast cell protease i	0.14817	0.24816	0.3188	46.1
Metallothionein ii	0.01257	0.19024	0.2222	56.4
Myod1	0.03577	0.10626	0.1982	54
Natriuretic	0.11933	0.15117	0.2103	54.2
Nucleolin	0	0.10577	0.246	43.3
Ornithine decarboxylase	0.01356	0.19883	0.2671	42.8
Oxytoxin-neurophysin i	0.01269	0.22051	0.1544	63.5
Peripherin	0.00559	0.14326	0.1944	54.8
Preprosomatostatin	0	0.09443	0.1519	47.4
Protamine 2	0.03588	0.09355	0.2407	60.6
Pulmonary surfactant protein sp-c	0.0432	0.20218	0.2138	48.4
Renin	0.09466	0.28986	0.2115	52.3
T-cell receptor v-b	0.0866	0.19231	0.1205	33.3
Thy-1 antigen	0.1072	0.18019	0.2495	51.4
Tnf receptor	0.02817	0.16922	0.2071	53.4
Ube1x	0	0.08754	0.1508	44
Vasopressin-neurophysin ii	0.0323	0.21009	0.1989	56
Interleukin 3	0.31367	0.08833	0.1925	47.5

[a] If all genes are considered using rank correlation, or if linear regression is used with *Interleukin 3* removed, a significant positive correlation is found between K_a and K_s ($P < 0.05$ and $P = 0.001$, respectively). No significant correlation is found between K_i and GC_i ($P = 0.639$). A one-sample t test on the difference between K_i and K_s reveals a significant difference ($P = 0.0047$), with K_i higher.

[b] K_a is the nonsynonymous exonic substitution rate, K_s is the synonymous exonic substitution rate, K_i is the intronic substitution rate, and GC_i is the intronic GC content. For accession numbers see HY97 (except for the *MyoD* rat gene, which should be M84176 rather than M84186). All values are means weighted by exon and intron size. Alignments were performed using the default settings of CLUSTALW. Tamura and Nei's (1993) multiple substitution correction method was used for all three K values, while Li's (1993) method was used to calculate K_a and K_s.

can make a difference. If, for example, we take the parametric analyses of the K_a/K_s covariation using the 14 alternative distance protocols, we can analyze the regression weighting each gene by its size. Doing so, we find that all correlations are now highly significant (Table 1), where previously only 4 of 14 were significant. One reason for this is that *Interleukin 3* being small is now less important as a data point. Comparing, then, the "weighted by gene" regressions with those unweighted for the data set minus the one outlier, we now find that both analyses agree that a highly significant correlation exists (Table 1).

Table 3 gives the K_a, K_s, K_i, and GC_i values for each gene using our recommended methods (with default CLUSTALW as the alignment protocol). These data, in combination with those for GC compositional correlations, contradict all four of HY97's findings.

Acknowledgments. We would like to thank Austin Hughes for his generous help and for the helpful comments from two anonymous referees. We are grateful to one of the referees for noting the problem with the use of β major and minor.

References

Altschul SF (1997) Sequence comparison and alignment. In: Bishop MJ, Rawlings CJ (eds) DNA and protein sequence analysis. IRL Press, Oxford, pp 137–168

Bernardi G, Mouchiroud D, Gautier C (1997) Isochores and synonymous substitutions in mammalian genes. In: Bishop MJ, Rawlings CJ (eds) DNA and protein sequence analysis. IRL Press, Oxford, pp 197–208

Comeron JM (1995) A method for estimating the numbers of synonymous and nonsynonymous substitutions per site. J Mol Evol 41: 1152–1159

Duret L, Mouchiroud D, Gautier C (1995) Statistical-analysis of vertebrate sequences reveals that long genes are scarce in GC-rich isochores. J Mol Evol 40:308–317

Genetics Computer Group (1994) Genetics Computer Group program manual for the Wisconsin Package, Version 8. Genetics Computer Group, 575 Science Drive, Madison, WI 53711

Hughes AL, Yeager M (1997) Comparative evolutionary rates of introns and exons in murine rodents. J Mol Evol 45:125–130

Ina Y (1995) New methods for estimating the numbers of synonymous and nonsynonymous substitutions. J Mol Evol 40:190–226

Jukes TH, Cantor CR (1969) Evolution of protein molecules. In: Munro HN (ed) Mammalian protein metabolism. Academic Press, New York, pp 21–123

Kimura M (1980) A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. J Mol Evol 16:111–120

Li WH (1993) Unbiased estimation of the rates of synonymous and nonsynonymous substitution. J Mol Evol 36:96–99

Li WH (1997) Molecular evolution. Sinauer Associates, Sunderland, MA

Li WH, Wu CI, Luo CC (1985) A new method for estimating synonymous and nonsynonymous rates of nucleotide substitution considering the relative likelihood of nucleotide and codon changes. Mol Biol Evol 2:150–174

Moriyama EN, Powell JR (1997) Synonymous substitution rates in *Drosophila:* mitochondrial versus nuclear genes. J Mol Evol 45: 378–391

Nei M, Gojobori T (1986) Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. Mol Biol Evol 3:418–426

Pamilo P, Bianchi NO (1993) Evolution of the *Zfx* and *Zfy* genes: rates and interdependence between the genes. Mol Biol Evol 10:271–281

Pesole G, Dellisanti G, Preparata G, Saccone C (1995) The importance of base composition in the correct assessment of genetic-distance. J Mol Evol 41:1124–1127

Rice P (19XX) Program manual for the EGCG Package. The Sanger Centre, Hinxton Hall, Cambridge CB10 1RQ, England

Tajima F, Nei M (1984) Estimation of evolutionary distance between nucleotide sequences. Mol Biol Evol 1:269–285

Tamura K (1992) Estimation of the number of nucleotide substitutions when there are strong transition-transversion and G + C content biases. Mol Biol Evol 10:512–526

Tamura K, Nei M (1993) Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. Mol Biol Evol 10:512–526

Chapter 11. Sex and conflict

Conflict between the levels of selection confounds the notion of adaptive optimality (see Chapter 1). Sexual reproduction increases the scope for conflict by enabling differences between replicators in transmission from one generation to the next, which promotes negative fitness covariance (Partridge and Hurst 1998).

Conflict is thought to have played an important role in the evolution of genetic systems (Hurst *et al.* 1992; Partridge and Hurst 1998). Conflict based explanations have been provided for syngamy (Hickey 1993; Hurst 1991b), mating strains and sexes (Hoekstra 1987; Hurst and Hamilton 1992), anisogamy and uniparental inheritance of cytoplasmic genes (Hastings 1992), and sex chromosomes and genomic imprinting (see Chapter 5). Of more relevance to the central theme of this thesis is the fact that conflict can affect genomic anatomy.

Copy number can be affected by conflict. The multicopy genes *Stellate* and *Suppressor of Stellate* on the X and Y chromosomes of *D. melanogaster* appear to be due to an ongoing conflict-based arms race (Hurst 1992b). Sex-specific copy number differences between sex chromosomes may result from conflict between the sexes (see Chapter 5). Similarly the sex-specific ploidy levels in haplodiploidy may be due to conflict between the maternal and paternal genomes (see Chapter 2). Effective copy number is reduced in a parent-of-origin specific manner in genomic imprinting, which may be due to conflict between mother and offspring (see Chapter 5). The enormous copy number of the *Alu* element within the human genome may be due to its selfish spread manifesting conflict within the nuclear genome (see Chapter 4).

The arrangement of replicators has important implications with regard to conflict. In the case of *Stellate* and *Suppressor of Stellate* the multiple copies are tightly linked. A number of selfish genes such as meiotic drive genes (Hurst 1999) and the *Medea* gene (see below and Research Paper 12) consist of two tightly linked elements: one coding for a toxin and one for the antidote. The transmission advantage of the selfish gene is dependent on tight linkage: the antidote is not advantageous unless toxin has been produced, and the toxin without the antidote is clearly disadvantageous.

In the rest of this chapter I shall not consider genomic anatomy, but rather the question of whether one can predict the consequences of conflict: is conflict ongoing or is it resolved in some way? It appears that the consequences of conflict are hard to predict (Hurst *et al.* 1996a). The evolution of genetic systems often appears to have set up "evolved constraints" which prevent one form of conflict, only for conflict to reappear in a new form (Hurst *et al.* 1996a). For example, the evolution of uniparental inheritance prevents the spread of fast replicating selfish cytoplasmic elements, but creates the grounds for cytoplasmic sex ratio distortion (Hurst 1993). The consequences of conflict are likely to be more extreme when the conflict is ongoing, which will be when power is shared evenly between the protagonists (Hurst 1993). When one side is much

stronger than the other, cooperation rather than conflict is predicted to prevail (Leigh 1991), as appears to be the case for meiotic drive genes which either go to fixation or are suppressed (Crow 1991). The balance of power will be dependent on sheer numbers and also biological details. In order to know whether a particular conflict is likely to lead to extreme consequences, we need to know whether the conflict is limited or unlimited. I have considered two cases of conflict, and have found that one selfish element appears to be self-limited, while another appears to be partially limited by suppression.

Reversible evolution in Medea

Werren *et al.* (1988) gave four reasons why selfish genes should be rare, in other words why conflict should be limited: negative frequency-dependent selection, counterbalancing negative effects upon individual fitness, evolution of suppressors and insensitives, and group selection against the selfish element. Hurst and McVean (1996) have suggested a special case which combines negative fitness effects and the evolution of insensitives, which is termed reversible evolution.

Reversible evolution can affect those selfish genes for which there are three alternative forms: killer (*K*), insensitive (*I*) and wildtype (*W*). *K* spreads when *W* is at high frequency as long as *K* is above a certain critical frequency, then *I* spreads when *K* is at high frequency, and finally *W* can reappear and spread to fixation when *I* is at high frequency. This set of intransitive allele relations is rather like the children's game of "rock", "scissors" and "paper". Scissors beats paper, then rock beats scissors, then rock loses to paper. The process of reversible evolution is rich in metaphorical implications: the selfish gene is hoist by its own petard, or, more biblically, it sows the seeds of its own destruction. In the context of the outcomes of conflict, reversible evolution ensures that conflict is self-limiting.

Medea is an acronym for Maternal-Effect Dominant Embryonic Arrest (Beeman *et al.* 1992). Mothers carrying *Medea* kill all their offspring who do not carry copies of *Medea* (Beeman *et al.* 1992). *Medea* spreads in a wildtype population as a result of the "fitness compensation" gained by the surviving members of the brood as the result of the death of their siblings. The *K* allele is the equivalent of *Medea* and is thought to express both a toxin and its antidote (tightly linked), the production of both of which is costly to individual fitness. The *W* allele is the wildtype which expresses neither toxin nor antidote, and the *I* allele expresses the antitoxin only. *Medea* genes appear to fulfil all the requirements of reversible evolution detailed above. *K* spreads in a *W* population due to offspring death and fitness compensation, but only when initially present above a certain threshold frequency. *I* can spread in a *K* population because it does not entail the cost of toxin production. Finally *W* can spread in a *I* population because it does not entail the cost of antitoxin production.

I examined the proposition that *Medea* is susceptible to reversible evolution by population genetics modelling (see Research Paper 12). I found that the intuitive explanation given above

does indeed hold, and that reversible evolution can occur subsequent to the invasion of a population by *Medea*. However, reversible evolution does not occur for all parameter values, and K can be maintained with the population reaching a stable equilibrium or a limit cycle. I concluded that the likelihood of reversible evolution is most strongly affected by the level of fitness compensation, because the costs of toxin and antitoxin production are likely to be low as the result of selection both on *Medea* and suppressors. High fitness compensation makes reversible evolution less likely because then the threshold frequency of K is then very low, which means that the population is unlikely to dip below the threshold.

The analysis of a more realistic model of *Medea*, in which the killing of offspring without *Medea* is incomplete, revealed an alternative reason for Medea to be self-limiting. With incomplete killing, *Medea* can become fixed even if there is a cost to producing toxin and antidote. Conflict comes to an end when *Medea* becomes fixed within a population, and can only reappear in a hybrid context, which is how the very first example of *Medea* was actually discovered.

Therefore *Medea* is self-limited in two ways: fixation within the population and reversible evolution. The self-limiting nature of *Medea* means that conflict is probably transient, and so in this case conflict is unlikely to have extreme consequences.

The evolution of early male-killers

Cytoplasmic factors tend to be inherited exclusively from mothers, and thus a cytoplasmic factor has zero fitness in a male (Hurst *et al.* 1997). Conflict arises because the optimal sex ratio for cytoplasmic genes (Cosmides and Tooby 1981) is more female biased than the optimal sex ratio for autosomal genes (Fisher 1930). A variety of cytoplasmic sex ratio distorting strategies are possible: parthenogenesis, feminisation and male-killing (see Hurst 1993). I have considered the evolutionary dynamics of those male-killers classified as "early" (sensu Hurst 1991a) (see Research Paper 13).

Early male-killers are bacteria which cause sex-specific pathogenesis to their insect hosts. The bacteria belong to a number of genera, most notably *Wolbachia*, *Rickettsia* and *Spiroplasma* (Werren *et al.* 1994; Werren *et al.* 1995; Williamson and Poulson 1979), and male-killing has been reported in a wide variety of insect taxa (Hurst *et al.* 1997; Hurst 1993). Early male-killing is thought to have evolved in male-killing bacteria as a consequence of the almost exclusively vertical (*i.e.* egg to brood) means of bacterial transmission between hosts (Hurst 1991a). Bacteria in male hosts are at an evolutionary dead-end, so male killing has a fitness cost of zero (from the bacterial point of view). But the death of males can augment the fitness of all remaining brood members, including female hosts carrying clonal relatives of the bacteria which killed males.

The option of early male-killing is not the only strategy remaining to a vertically transmitted bacterium . If vertical transmission is inescapable, then mutually beneficial symbiosis, cooperation rather than conflict, may provide an evolutionary solution for the bacterium (Douglas 1989). On the other hand, if the bacterium can evolve horizontal transmission then it may become

more like a conventional pathogen by seeking to maximise its rate of spread through the population (Hurst *et al.* 1997).

This positive effect of male-killing on surviving brood members is termed fitness compensation, and can occur for a number of reasons: reduced intra-brood competition, reduced inbreeding or direct benefits from egg cannibalism (for references see Hurst *et al.* 1997). As with *Medea*, it is fitness compensation which allows male-killers to spread in a host population, despite infected male hosts being killed and infected female hosts bearing a fitness cost (Hurst 1991a).

It is a property, and indeed a definition, of all selfish elements that their spread creates the context for the spread of suppressor genes (Hurst *et al.* 1996a). I considered the evolutionary dynamics of a population containing two different male killers as well as hosts which could evolve resistance against one of the male-killers. The original focus of the research was to see if host resistance could explain the male-killer polymorphisms observed in nature (Hurst *et al.* 1999), since such empirical findings appeared to be at odds with simple models of male-killing which invoke no host resistance. I confirmed that host resistance provides a potential explanation for male-killer polymorphism and the maintenance of conflict, but I also found reasons why conflict between male-killers and their host is likely to be transient.

Competition between two strains of male-killers within a population is determined by the Basic Rate of Increase (*BRI*), which is a function of fitness compensation, the vertical transmission rate, and cost to the host of male-killer possession independent of male-killing. The male-killer with the highest *BRI* always wins, which means that in the absence of host resistance the population will possess a single male-killer. This male-killer will have a high vertical transmission rate, will incur low costs on its hosts independent of male-killing, and will be present at a high frequency. This situation reveals an important difference between meiotic drive genes and male-killers. When meiotic drivers are at high frequencies there is little conflict, and the cost to the population is minimal. But when male-killers are at a high frequency, almost all the males in the population are killed, and consequently the population suffers a high cost. If a male-killer with complete killing ever reached fixation (this requires perfect vertical transmission), no males would be produced and the population would go extinct.

Without host resistance one expects a ratchet-like mechanism, whereby the *BRI* of male-killers increases over time, which means that the equilibrium frequency of male-killers, and the level of male-killing, also tends to increase. The ratchet may be restricted to some extent by a positive correlation between male-killer costs and transmission rates, but the existence of very high frequencies of male-killers in some populations (Majerus *et al.* 1998) indicates that such limitations are probably weak. In contrast, it appears that the evolution of host resistance can counter the ratchet in two ways. Firstly, the spread of a costless or highly effective resistance gene can drive the male-killer from the population. Thus the ratchet can be reset. Secondly, the evolution of host resistance to an incumbent high *BRI* male-killer allows a lower *BRI* male-killer unaffected by host resistance to displace the higher *BRI* male-killer. Thus the ratchet can be reversed. The alternative

result following the invasion of a lower *BRI* male-killer unaffected by host resistance is a multiple male-killer polymorphism, which does not counteract the ratchet. However multiple male-killer polymorphism is unlikely unless the viability cost of host resistance is greater than about 1%.

Thus the conflict between male-killers and their hosts is partially limited by the evolution of host resistance. However the conflict does seem to be ongoing even if the level of conflict does not necessarily tend to increase over time. So it seems likely that this conflict could lead to further consequences. Arms races might take place between host resistance genes and counter-resistance bacterial genes, perhaps promoting the evolution of multiple copy numbers to increase dosage and linkage between the multiple copies to allow control of expression.

Research Paper 12. The dynamics of maternal-effect selfish genetic elements

Nick Smith (1998)
J. theor. Biol. **191** 173-180.

J. theor. Biol. (1998) **191**, 173–180

The Dynamics of Maternal-effect Selfish Genetic Elements

NICK G. C. SMITH

Centre for Mathematical Biology, Department of Biology and Biochemistry, University of Bath, Claverton Down, Bath BA2 7AY, U.K.

(*Received on* 7 *July* 1997, *Accepted on* 28 *October* 1997)

Maternal-effect selfish genes such as *Medea* or *Scat* act to kill progeny that do not bear a copy of the selfish gene present in the mother. Previous models of this system allowed for two types of allele, the selfish (killer) type and the sensitive (susceptible) wild-type. These models predict that the invasion conditions of the selfish allele are quite broad and that if invasion is possible a high frequency equilibrium is to be expected. The selfish element is therefore predicted to persist. Here a hypothetical third allele that neither kills nor is killed (i.e. insensitive) is considered. Such an allele could enter a population by recombination, mutation or migration. The incorporation of this third allele profoundly affects the dynamics of the system and, under some parameter values, it is possible for the spread of the insensitive allele to lead, eventually, to the fixation of the wild-type allele (reversible evolution). This is most likely if the death of progeny provides no direct benefit to the surviving sibs (i.e. in the absence of fitness compensation), as in insects without gregarious broods. Under these circumstances the selfish element cannot spread when infinitely rare, only after having risen to some finite frequency. A fitness cost to bearing the killer allele then causes its loss. However, if fitness compensation is found (e.g. in placental mammals) the invasion of the selfish element from an infinitely low level is possible for a wide range of costs and both stable coexistences of all three alleles and limit cycles of all three are then found. It is therefore to be expected that in mammals selfish maternal-effect genes are more likely both to spread and to persist than in insects, due to their different levels of fitness compensation.

1. Introduction

Selfish genetic elements (SGEs) enhance their own transmission relative to the rest of the genome of the individual in which they appear and as a consequence they may spread deterministically in a population even if either neutral or detrimental to the organism as a whole (Hurst & McVean, 1996; Werren *et al.*, 1988). Given that they often have relatively broad invasion conditions, it is perhaps curious that selfish elements are not more common (but see Jaenike, 1996). It is then pertinent to ask under what circumstances selfish elements will not be held in polymorphism. Several possibilities have been suggested (for mechanisms limiting the spread of SGEs see Werren *et al.*, 1988). In some instances, the element may go to fixation, either sending the population extinct or effectively disappearing (only to potentially reappear in hybrids which was how maternal-effect selfish genes were first discovered (Beeman *et al.*, 1992)). Alternatively unlinked suppressors may evolve that eliminate the element from the population. Both mechanisms have been suggested as means by which Mendelian segregation is maintained (Crow, 1991).

Recent analysis of the dynamics of cytoplasmic incompatibility suggests a further possibility in which, in the absence of unlinked suppressors, a population may cyclically evolve from fixation of wild-type back to fixation of wild-type after invasion of a selfish element (Hurst & McVean, 1996). These authors consider a system with three cytotypes: killers, insensitives and wild-types. The invasion of distorting killer elements allows for the spread of insensitive elements which have no distorting effect themselves but which confer immunity to the action of the distorter. Once insensitives are common the original wild-type sensitive allele can spread to fixation. The

0022–5193/98/060173 + 08 $25.00/0/jt970579

trajectory, starting from fixation of wild-type and eventually decaying back to fixation of wild-type is described as "reversible evolution". Reversible evolution is not a general property of all systems of selfish elements with three alleles ("killer", insensitive and sensitive) (see e.g. Charlesworth & Hartl, 1978; Nauta & Hoekstra, 1993).

The fundamental difference between meiotic drive and cytoplasmic incompatibility which allows reversible evolution in the latter but not the former lies in their differing abilities to spread from infinitely low levels despite being costly to individuals bearing them. Meiotic drive SGEs spread by over-representation in the gametes passed on to the next generation, and can hence spread from infinitely rare as long as the destruction of 50% of a heterozygote male's sperm causes a fertility reduction of less than 50%. In contrast cytoplasmic incompatibility SGEs spread by reducing the progeny of those crosses which would yield uninfected offspring, and if the element is infinitely rare then infinitely few crosses give reduced progeny and the relative advantage to the element is correspondingly small, which means that cytoplasmic elements can only spread above a critical frequency if they are costly (Hurst & McVean, 1996). If a SGE can spread from infinitely low levels then reversible evolution is not possible but if the distorter can only spread above a certain level then if its frequency drops below this level due to successive spreads of insensitives and wild-types it will be unable to recover and will be lost to the population (see Fig. 1).

Hurst & McVean (1996) suggest that, due to maternal-effect selfish elements and cytoplasmic incompatibility both spreading by the death of relatives who do not share a copy of the selfish gene, maternal-effect SGEs might also undergo reversible evolution. Here the theoretical population dynamics of maternal-effect selfish elements are examined to determine whether reversible evolution is possible for such elements and, if so, under what conditions.

2. Maternal-effect Selfish Genes

Mothers carrying maternal-effect selfish genes kill all their offspring who do not carry copies of these genes. The selfish gene enhances its transmission relative to its competitors by causing the death of individuals bearing sensitive alleles. The killing clearly lowers the mother's fitness as the number of offspring produced falls.

Three examples, supported by segregation data, have been described. The *Medea* gene in the flour beetle *Tribolium castaneum* was the first maternal-effect selfish gene to be described (Beeman *et al.*, 1992). Its presence was inferred from reciprocal hybrid female semi-sterility between geographically diverse strains. The means of inheritance of murine severe combined anaemia and thrombocytopenia (*Scat*) was demonstrate by Peters & Barker (1993), and the equivalence of *Scat* and *Medea* was pointed out by Hurst (1993). Another case of maternal-effect selfish genes in mice was recently demonstrated by Weichenhan *et al.* (1996). One high copy long-range repeat cluster *Mut* is found on chromosome 1. Preferential recovery of maternal *Mut* in female *Mut*/+ and male +/+ crosses was shown to be due,

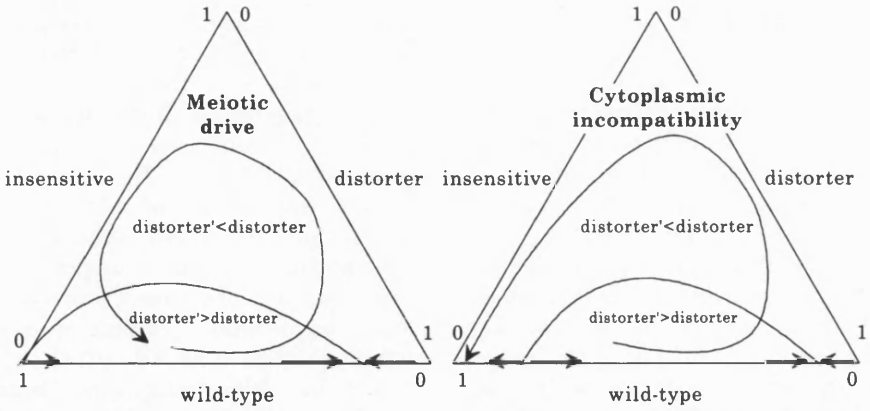


FIG. 1. The reason for reversible evolution affecting cytoplasmic incompatibility but not meiotic drive is illustrated graphically in allele space showing the frequencies of the wild-type, distorter, and insensitive alleles. The meiotic drive distorter allele is able to spread from infinitely rare, hence the line of distorter′ = distorter intersects the wild-type axis at 0. But the cytoplasmic incompatibility distorter allele is only able to spread above a critical frequency which is the wild-type axis intercept with the distorter′ = distorter line. For both types of element the effect of the sequential spread of distorter then insensitive then wild-type is shown. While the meiotic drive distorter allele can always spread however close the population gets to wild-type fixation, the cytoplasmic incompatibility wild-type allele can reach fixation without crossing the distorter′ = distorter line.

at least in part, to lethality of post-implantation +/+ embryos. *Mut* differs from *Scat* in genomic location (chromosome 1 as opposed to 8) and in being found in wild populations of *Mus musculus domesticus*, rather than laboratory strains of mice.

3. The Model

Wade & Beeman (1994) have analysed the population genetics of *Medea* considering only a sensitive and a killer gene, and predicted fixation of killers unless killers had associated costs in which case a stable killer/sensitive polymorphism would result. Here it is asked what would happen were a third allele to be introduced, namely an insensitive, which does not kill progeny and which is not affected by the killer allele?

In the case of maternal effect selfish elements, such an allele remains hypothetical. However, assuming that like other selfish elements, such as meiotic drive genes (e.g. *Segregation distorter* of *Drosophila melanogaster* or *t-complex* of mice) the selfish allele is actually two very tightly linked alleles, one producing a toxin and one the antitoxin, the possibility of an insensitive allele/chromosome (cf. $Sd^{+}Rsp^{i}$ of *Drosophila*) seems quite plausible and its dynamics worth investigating. That *Scat* might be two loci has been suggested (Hurst, 1993) as two locus distorters typically are found near centromeres (due to the need for reduced recombination) and *Scat* is very close to the centromere on mouse chromosome 8. Similarly *Mut* is contained in a region of low recombination and two types of allele have been inferred in this instance (see Agulnik *et al.*, 1993).

Consider then an infinite population of diploid sexual organisms. There is a single autosomal locus for the selfish gene, and three alleles *m*, *w*, and *i*. The alleles can be considered as differing in production of toxin and antitoxin (see above). The *m* allele codes for both toxin and antitoxin (from a two-allele viewpoint the distorter and insensitive alleles are linked), *i* for antitoxin only (non-distorter and sensitive alleles linked in a two-allele view), and *w* is the wild-type null allele (neither toxin nor anti-toxin is produced, in two-allele terms non-distorter and sensitive alleles are linked).

Maternal-effect killing occurs if a toxin-producing mother has progeny without a copy of the antitoxin-coding allele (of either maternal or paternal descent). One copy of an antitoxin-bearing allele is all that is required for complete rescue from maternal-effect killing. There is a single father per brood. All sensitive progeny die. The three alleles incur different viability costs due to production of toxin and antitoxin, U_t and U_{at}, respectively. Random mating is assumed, but population substructure can lead to fitness compensation: the remaining offspring within a brood gain an advantage when some of their sibs have been killed. The fitness advantage to each remaining brood member is G_1 when one-quarter of the brood have been killed, and G_2 when one-half have been killed. Since those families in which maternal-effect killing takes place will not be fitter than families with no maternal-effect deaths, then the upper bounds of G_1 and G_2 are 0.33 and 1.0, respectively. These "gain" values will decrease as local density regulation decreases either because the mother does not supply the same resources to fewer offspring, or analogously because broods are not fully separated so non-family members also benefit. This model thus incorporates what is termed "family-level soft selection" by Wade (1985). Wade & Beeman (1994) showed that while intra-family selection always favours selfish maternal-effect elements, inter-family selection always acts against maternal-effect SGEs. The greater the reduction of the gain values below their upper bounds the stronger the inter-family selection relative to intra-family selection.

For the one locus model considered here to be formally analagous to a two locus two allele model the remaining fourth possible allele must be considered. This allele codes for toxin but not antitoxin and is similar to a meiotic drive "suicide" chromosome. Heterozygotes with one suicide allele can survive but suicide homozygotes are always killed (the mother produces toxin and the offspring possess no antitoxin). Simulations of the four allele one locus model were carried out (using a mutation model, see Section 4.2), and gave results qualitatively identical to those of the three allele one locus model (for the same parameters both models predicted the same outcome). The suicide allele never rose above very low levels. Hence only the simpler three allele one locus model is discussed.

From the model a set of recursive equations is obtained (see the Appendix: *d* and *k* are parameters for incomplete rescue and incomplete killing, respectively, and for the model described above take values of 0 and 1, respectively). The invasion conditions for the two allele systems are tractable to algebraic analysis. But problems such as initial levels required for non-deterministic spread, two allele stability analysis, and any analysis of the three allele system are all too complex to solve fully by analysis. This is because the maternal-effect killing violates Hardy–Weinberg and so the recursions are for genotype rather than allele frequencies. So to study the dynamics of the model here computer simulations

have been used. Visualising the simulation results requires a simplification of 6-D genotype space to 3-D allele space, which can be presented in a 2-D plot since the allele frequencies sum to 1. It is important to remember when viewing plots of allele space that each point conceals an infinity of possible genotype distributions.

4. Results

4.1. INVASION OF THE SELFISH GENE

The invasion of w by m has to be the first part of any putative cycle, and this part of the investigation mirrors Wade & Beeman's (1994) look at the population dynamics of *Medea*. The conditions for the deterministic invasion of w by m from infinitely low levels are given by the solutions to $\delta mw'/\delta mw > 1$ at $m = 0$ and $w = 1$. These are that $(1 - U_{at})(1 - U_t)(1 + G_2/2) > 1$. This result agrees with that of Wade & Beeman (1994).

To ease the investigation of the behaviour of the system it is helpful to make the parameter simplifications of $U = U_t = U_{at}$ and $G = G_2 = 2\ G_1$, and so the requirement for invasion becomes $(1 - U)^2 > 1/(1 + G/2)$.

Spread of maternal-effect lethals in populations is possible outside of this parameter range. A similar result has been described for cytoplasmic incompatibility (Hurst & McVean, 1996). For both cytoplasmic incompatibility and maternal-effect killing the rate of spread of the selfish gene is dependent on its frequency: initially the more of them there are the greater the probability that wild-types will die. Hence, in both systems, if the costs are too high for deterministic invasion when infinitely rare then the selfish allele can nonetheless still invade, provided its initial level is high enough. Stochastic effects, small population sizes and structured populations are likely to provide the necessary conditions (Bull *et al.*, 1992). The required level of initial m increases quite sharply into the region of several percent once the costs of producing toxin and antitoxin are only about 0.01 (in terms of selective coefficients) above those which allow deterministic invasion (see Fig. 2).

Given that conditions for the spread of m are broad, we can now ask about possible equilibria positions. The conditions for the deterministic invasion of w into a population of m are given by the solutions to $dmw'/dmw > 1$ at $w = 0$ and $m = 1$. With the parameter simplifications this resolves to $(1 - U)^2 < 1$, which holds for $U > 0$. Thus if m can invade a m/w equilibrium will result unless both U_t and U_{at} are zero in which case m will become fixed. The equilibrium proportion of w increases as U_t, U_{at} and G_1 increase (U_t and U_{at} because then mm suffers more relative to mw, and G_1 because mw progeny of mw/mw crosses are more favoured). These results again support those of Wade & Beeman (1994).

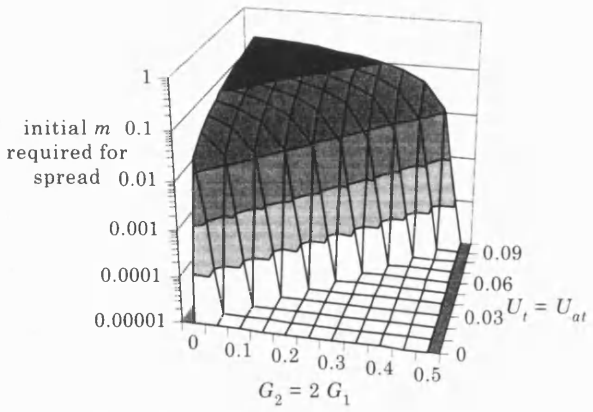


FIG. 2. The initial frequency of the m allele required for non-deterministic spread is plotted against simplified parameter space ($G_2 = 2\ G_1$ and $U_t = U_{at}$). Deterministic spread is given a default level of 0.00001. The plot shows that beyond deterministic spread the initial required level of m increases sharply.

4.2. SPREAD OF THE INSENSITIVE ALLELE

To investigate the hypothetical three allele system, i must be introduced. There are three ways to introduce i which are expected to give very similar results: migration (invasion model), mutation (mutation model) and recombination (two-allele two-locus model). A recombination model was not developed, partly because such segregation distorters tend to be found in regions where recombination is suppressed, for example, *Mut* incorporates a paracentric inversion (Weichenhan *et al.*, 1996) and *Scat* is linked to the centromere.

The invasion model allowed the population to reach m/w equilibrium and then a variable proportion of i was added to the population, in the form of ii individuals. The recursive equations would then operate for all six genotypes once Hardy–Weinberg reassortment had spread i evenly through the genotypes before the next generation.

The mutation model envisaged i alleles produced by a null mutation to the toxin part of the m allele (at rate *toxm*). Similarly the antitoxin part of the i allele could mutate to give a w allele (at rate *atoxm*). The generation of a "suicide" allele by the mutation of the antitoxin part of m was ignored, as were back mutations.

For both the invasion and mutation models spread of i was trivially dependent on $U_t > 0$. This result is the solution to $dmi'/dmi > 1$ at $m = 1$ and $i = 0$. The system kinetics of the two models at the same U and

G values were very similar, which agreed with Nauta & Hoekstra's (1993) assertion that the method of resistant introduction did not matter, and augured well for conclusions that did not depend on the method of i introduction.

4.3. THREE ALLELE DYNAMICS

From the above we may conclude that m can reach equilibrium with w from low initial levels, and then, under broad realistic conditions, i can spread from low levels. What will then happen?

For both the invasion and mutation parameter scans, four different outcomes were observed (see Fig. 3 for a visualisation of how the results change over G-U parameter space for a run of the mutation model). First, m could fail to spread initially. Second, reversible evolution could occur, i.e. m spreads only to be replaced by i then w becomes fixed. Third, a stable limit cycle could be entered. Finally, the system could come to rest at an internal equilibrium (for plots of the three possible outcomes once m invades see Fig. 4). When costs are low, the system moves slowly and internal equilibria may be mis-classified as limit cycles. But investigations have shown the presence of stable limit cycles at certain parameter values. Again the invasion and mutation models give nearly identical results (one difference being that the invasion model could give $m = 1$ as a stable solution, but m could never become fixed with the mutation model). Simulations show that the outcome of a run depends mainly on the G and U values, but also slightly on model parameters such as the amount of *ii* introduced in the invasion model and the mutation rates in the mutation model. Hence I shall provide a fairly qualitative description of the dynamics of the model.

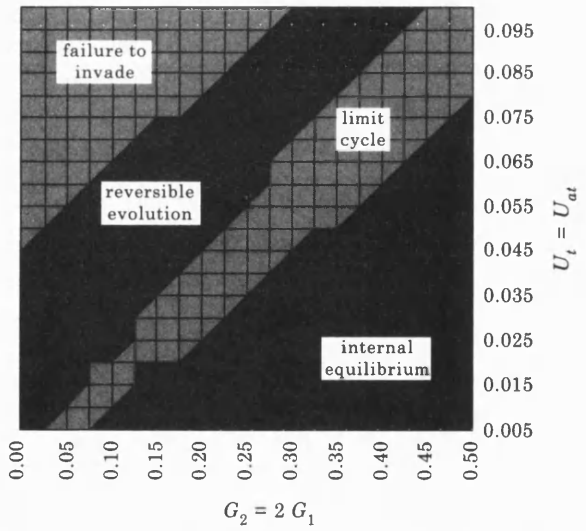


FIG. 3. The different outcomes of the mutational model ($m(0) = 0.1$, $atoxm = toxm = 10^{-7}$ and limit cycle/internal equilibrium decision made after 100000 generations) across simplified parameter space ($G_2 = 2\ G_1$ and $U_t = U_{at}$). As gains increase from low levels with costs held steady the outcomes change as follows: failure of the m allele to spread, population returns to $w = 1$; reversible evolution with $w = 1$ as the end result; a limit cycle with all six genotype frequencies continuously changing; and finally a stationary equilibrium of all six genotype frequencies. If the initial level of m is reduced then the width of the reversible evolution band decreases.

So why do the outcomes change as G-U parameter space is traversed?

(i) Reversible evolution or limit cycle?

Reversible evolution requires that m can only spread from finite levels (see Introduction). But if the finite level required is low a limit cycle rather than reversible evolution occurs (since the value of m may never fall low enough for m to be unable to spread). With limit cycles the system goes close to boundaries, and elimination of some alleles by chance may be likely (cf. Charlesworth & Hartl, 1978). However elimination of rare alleles failed to make reversible evolution more likely because i was usually lost before m.

(ii) Limit cycle or internal equilibrium?

Internal equilibria result when G is well in excess of the level needed for deterministic spread of m in a population of w. If the gains are much greater than the costs, then w cannot ever spread to high levels and provide the conditions for m to spread again. Hence the system finds an equilibrium. Limit cycles are restricted to a fairly narrow band of gains above the critical costs.

4.4. RELAXING THE COST AND GAIN PARAMETER ASSUMPTIONS

If the assumptions of $G_2 = 2\ G_1$ and $U_t = U_{at}$ are relaxed, the proportions of parameter space leading to the various possible outcomes change significantly.

(a) G_2 not equal to $2\ G_1$

It seems a reasonable assumption that the benefit to the remaining sibs is doubled when the proportion killed is doubled. The relative increases in the proportion of total resources (for G_1 from $\frac{1}{4}$ to $\frac{1}{3}$, and for G_2 from $\frac{1}{4}$ to $\frac{1}{2}$) suggest $G_2 > 2\ G_1$, while diminishing returns suggest that G_1 should be larger relative to G_2. An increase in G_1 relative to G_2 makes limit cycles become more likely relative to reversible evolution, and internal equilibria more likely relative to limit cycles. G_2 helps m only (see the Appendix, G_2 only appears in crosses between ww males and mw females) but G_1 helps both m and i, and thus high levels of G_1 make the stable maintenance of both i and m more likely.

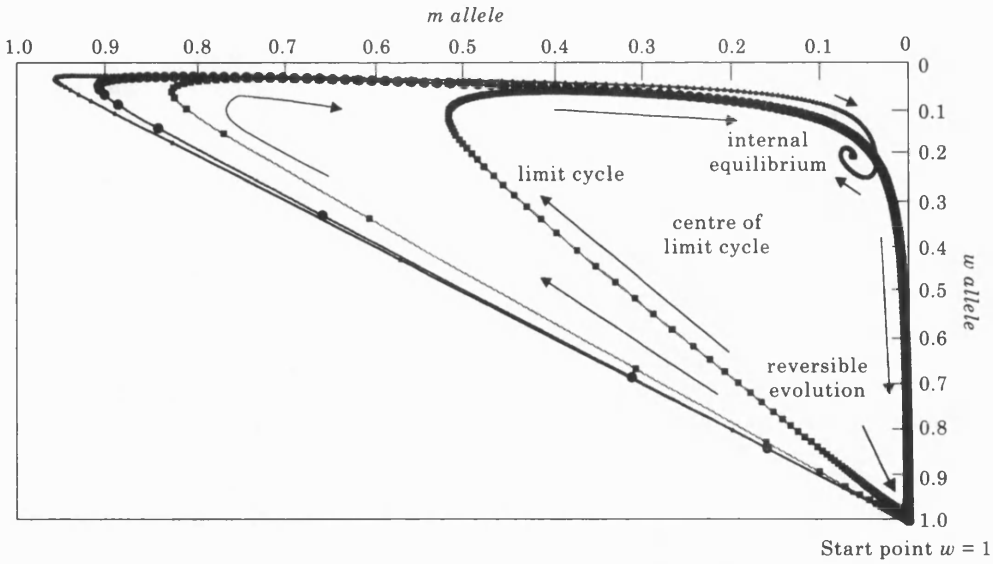


FIG. 4. Three different outcomes of the mutation model at different cost and gain parameter values ($U = U_t = U_{at}$ and $G = G_2 = 2\ G_1$) are plotted on simplified genotype space (only allele frequencies shown and hence an infinite number of genotype distributions contained within each point; m and w can be directly read off axes while i is a function of the perpendicular distance from the line between $w = 1$ and $m = 1$). The reversible evolution plot starts at finite m (0.1) since deterministic spread is not possible under such parameter values ($U = 0.1$, $G = 0.4$) while both the limit cycle ($U = 0.1$, $G = 0.5$) and internal equilibrium ($U = 0.07$, $G = 0.5$) plots start at very low m (0.001). All three plots move in a clockwise direction since m spreads in w, i spreads in m, and w spreads in i. The internal equilibrium plot gets stuck at a high i and low m and w position. Both the limit cycle and reversible evolution plots approach $w = 1$. While the limit cycle parameters enable deterministic spread of m in a w population and allow m to increase and start the cycle again, the reversible evolution parameters make $w = 1$ a stable attractor if m falls below a certain threshold (in this case about 0.04). The reversible evolution plot falls below this threshold and thus w is fixed. Key: —■— limit cycle; —●— reversible evolution; —▲— internal equilibrium.

(b) U_t not equal to U_{at}

There is no reason for the cost of producing a toxin to equal the cost of producing the respective antitoxin, and no basis for such an assumption given that no mechanism of maternal lethality is known. So U_t/U_{at} was varied for a constant C ($C = (1 - U_t)(1 - U_{at})$) which, as $G_2 = 2\ G_1$ varied, would go through parameter sets that had given failure to spread, reversible evolution, limit cycle and internal equilibrium with $U_t = U_{at}$.

When $U_{at} \gg U_t$ the only possible solutions are failure to spread and internal equilibrium, and as U_{at}/U_t decreases until $U_t \gg U_{at}$ both reversible evolution and limit cycles become more likely. Both i and m suffer from the cost of U_{at} and the difference in costs is determined by U_t. Thus if m can invade initially when $U_{at} \gg U_t$, then i is unable to spread to a high level since its advantage relative to m is slight. So w cannot reach a high level by outcompeting i. Hence the population can never be sucked into the domain where fixation of w acts as an attractor. On the other hand, if U_t is high and U_{at} low then i enjoys a huge advantage over m and takes over the population to a large extent, leaving itself vulnerable to being ousted by w. Since m is reduced to a very low level by the spread of i it will be unable to spread in a population of w.

These results show that $U_t \gg U_{at}$ favours reversible evolution, and so any estimations of reversible evolution probability ought to state the assumed relationship.

5. Discussion

The results in Section 4.1 agree with the work of Wade & Beeman (1994) in showing that the invasion of maternal-effect selfish elements is possible under a broad range of parameter values and that the m and w alleles achieve equilibrium over a wide range of parameters. So to what extent is reversible evolution a convincing explanation for why only three cases of maternal-effect selfish element polymorphisms are known?

How different levels of gains and costs affect the likelihood of reversible evolution is shown by Fig. 3. If costs are low and gains are high then reversible evolution is not to be expected with limit cycles and internal equilibria being more likely outcomes. But if costs are high relative to gains, then if m can spread initially, reversible evolution is the probable outcome.

A parallel can be drawn here with the dynamics of 'male killers' (Hurst, 1991). Fitness compensation is required for the spread and maintenance of male

killers, just as high fitness compensation predisposes maternal-effect selfish elements to be maintained in a population rather than suffer from reversible evolution.

Relaxing some of the simplifying assumptions causes important qualitative changes in the dynamics of the system. If G_1 is higher than $G_2/2$ reversible evolution is less likely (relative to Fig. 3), and conversely, if it is lower then reversible evolution is more likely. U_t not equal to U_{at} and unequal toxin costs in the sexes (expected on the basis that only mothers kill) both cause similar changes. If $U_{at} > U_t$ (equivalent to males bearing smaller toxin costs) reversible evolution is less likely, and if $U_t > U_{at}$ then reversible evolution is more likely.

Given these findings what do the data from different organisms suggest as to the chances of reversible evolution for maternal-effect SGEs? Consider the basic parameters of cost (of producing toxin and/or antitoxin) and gain (fitness compensation) because it is these parameters which most affect the outcome of *m* invasion. Costs appear to be low (Beeman *et al.*, 1992), though this is hard to determine in *Scat* since almost all organisms not bearing selfish genes suffer from maternal-effect lethality. A lack of a molecular understanding of the processes involved makes predicting costs impossible, and indeed some mechanisms would incur no costs at all (for some possibilities see Weichenhan *et al.*, 1996). Whatever the mechanism it is probably fair to argue that costs are likely to have been minimised by selection on selfish genes themselves or else on modifiers, and hence the level of gains is the more useful indicator of the chances of reversible evolution.

In general the level of gains is dependent on population sub-structure, the timing of offspring killing, and parental resource control. In flour beetles the killing occurs at a fairly late stage (Beeman *et al.*, 1992), hence low fitness compensation is expected. McCauley & Wade (1980) have shown that under controlled laboratory conditions *T. castaneum* strains do show local density regulation within families. Thus fitness compensation is expected to be fairly low in insects. Extended parental care means that gains are likely to be high in mammals. With *Mut* the recovery ratios are distorted after implantation, which would make gains large unless the mother decreases parental investment as a response to embryo reabsorption.

So insects are predicted to have low fitness compensation gains and thus reversible evolution appears likely. However, for mammals with their high gains reversible evolution does not appear to offer a viable solution to the problem of the paucity of maternal-effect selfish elements.

How then is it possible to explain the lack of such elements in mammals? The difficulties of large breeding experiments with mammals mean that studies sufficiently large to show maternal-lethality effects must be rare, even if experimenters are looking for them. Perhaps maternal-effect selfish elements only get created occasionally (especially likely if simultaneous production of a toxin and an antitoxin is required), and thus the rarity of maternal-effect elements would be simply due to the low rate of creation rather than difficulty of maintenance.

The main model presented here carries several simplifying assumptions, including complete killing and complete rescue. In *Scat* not all *Scat*/*Scat* (*ww*) offspring of *Scat*/+ (*mw*) mothers are killed, around 7.5% entering an extended remission period during which some can produce young (Peters & Barker, 1993). Thus killing is not always complete (the surviving proportion is considered to be $1 - k$ in the table of crosses in the Appendix). Incomplete rescue (e.g. in flour beetles, see Beeman *et al.*, 1992) is asymmetric with regard to sex of parents and will therefore have a more complex effect than simply increasing the cost of bearing a toxin allele. In the Appendix the proportion of offspring killed is *d* times the number of maternal copies of toxin divided by the number of offspring copies of antitoxin. Analysis shows that incomplete rescue can lead to the fixation of *m* due to *mw* suffering more from incomplete rescue than *mm* due to fewer antitoxin copies (solving $\delta mw'/\delta mw > 1$ for $m = 1$ and $w = 0$ gives $(1 - U_t)(1 - U_{at}) < (1 - 3d/2)/(1 - d)$ as the requirement for *w* spread in *m*, and clearly for $2/3 > d > 0$ there will always be some value of $1 > U_t = U_{at} > 0$ below which *m* will be fixed). This result of a more realistic model provides an alternative explanation to reversible evolution for why maternal-effect elements are not more commonly found in polymorphism. For mammals this explanation is appealing because the expected high levels of fitness compensation make fixation possible even for low levels of incomplete rescue. Furthermore simulation results suggest that both incomplete rescue and complete rescue increase the proportion of parameter space which yields reversible evolution, as long as the selfish element is able to spread initially.

For advice, support and encouragement thanks to Laurence Hurst, Steve Freeland, Mike Mogie, and the Centre for Mathematical Biology at Bath University.

REFERENCES

AGULNIK, S. I., SABANTSEV, I. D., ORLOVA, G. V. & RUVINSKY, A. O. (1993). Meiotic drive on aberrant chromosome 1 in the mouse is determined by a linked distorter. *Genet. Res.* **61**, 91–96.

BEEMAN, R. W., FRIESEN, K. S. & DENELL, R. E. (1992). Maternal-effect selfish genes in flour beetles. *Science* **256**, 89–92.

BULL, J. J., MOLINEUX, I. J. & WERREN, J. H. (1992). Selfish genes. *Science* **256**, 65–65.

CHARLESWORTH, B. & HARTL, D. L. (1978). Population dynamics of the segregation distorter polymorphism of *Drosophila melanogaster*. *Genetics* **89**, 171–192.

CROW, J. F. (1991). Why is mendelian segregation so exact. *Bioessays* **13**, 305–312.

HURST, L. D. (1991). The incidences and evolution of cytoplasmic male killers. *Proc. Roy. Soc. Lond. B* **244**, 91–99.

HURST, L. D. (1993). *Scat+* is a selfish gene analogous to *Medea* of *Tribolium castaneum*. *Cell* **75**, 407–408.

HURST, L. D. & MCVEAN, G. T. (1996). Clade selection, reversible evolution and the persistence of selfish elements—the evolutionary dynamics of cytoplasmic incompatibility. *Proc. Roy. Soc. Lond. B* **263**, 97–104.

JAENIKE, J. (1996). Sex-ratio meiotic drive in the *Drosophila quinaria* group. *Am. Nat.* **148**, 237–254.

MCCAULEY, D. E. & WADE, M. J. (1980). Group selection: the genetic and demographic basis for the phenotypic differentiation of small populations. *Evolution* **34**, 813–821.

NAUTA, M. J. & HOEKSTRA, R. F. (1993). Evolutionary dynamics of spore killers. *Genetics* **135**, 923–930.

PETERS, L. L. & BARKER, J. E. (1993). Novel inheritance of the murine severe combined anemia and thrombocytopenia (*Scat*) phenotype. *Cell* **74**, 135–142.

WADE, M. J. (1985). Soft selection, hard selection, kin selection, and group selection. *Am. Nat.* **125**, 61–73.

WADE, M. J. & BEEMAN, R. W. (1994). The population-dynamics of maternal-effect selfish genes. *Genetics* **138**, 1309–1314.

WEICHENHAN, D., TRAUT, W., KUNZE, B. & WINKING, H. (1996). Distortion of mendelian recovery ratio for a mouse hsr is caused by maternal and zygotic effects. *Genet. Res.* **68**, 125.

WERREN, J. H., NUR, U. & WU, C.-I. (1988). Selfish genetic elements. *Trends Ecol. Evol.* **3**, 297–302.

APPENDIX

Father	Mother	*mm*	*mi*	*mw*	*ii*	*iw*	*ww*
mm	*mm*	$1(1-d)$					
mi	*mm*	$0.5(1-d)$	$0.5(1-d)$				
mw	*mm*	$0.5(1-d)$		$0.5(1-2d)$			
ii	*mm*		$1(1-d)$				
iw	*mm*		$0.5(1-d)$	$0.5(1-2d)$			
ww	*mm*			$1(1-2d)$			
mm	*mi*	$0.5(1-d/2)$	$0.5(1-d/2)$				
mi	*mi*	$0.25(1-d/2)$	$0.5(1-d/2)$		$0.25(1-d/2)$		
mw	*mi*	$0.25(1-d/2)$	$0.25(1-d/2)$	$0.25(1-d)$		$0.25(1-d)$	
ii	*mi*		$0.5(1-d/2)$		$0.5(1-d/2)$		
iw	*mi*		$0.25(1-d/2)$	$0.25(1-d)$	$0.25(1-d/2)$	$0.25(1-d)$	
ww	*mi*			$0.5(1-d)$		$0.5(1-d)$	
mm	*mw*	$0.5(1-d/2)$		$0.5(1-d)$			
mi	*mw*	$0.25(1-d/2)$	$0.25(1-d/2)$	$0.25(1-d)$		$0.25(1-d)$	
mw	*mw*	$0.25(1+G_1.k)(1-d/2)$		$0.5(1+G_1.k)(1-d)$			$0.25(1-k)$
ii	*mw*		$0.5(1-d/2)$			$0.5(1-d)$	
iw	*mw*		$0.25(1+G_1.k)(1-d/2)$	$0.25(1+G_1.k)(1-d)$		$0.25(1+G_1.k)(1-d)$	$0.25(1-k)$
ww	*mw*			$0.5(1+G_2.k)(1-d)$			$0.5(1-k)$
mm	*ii*		1				
mi	*ii*		0.5		0.5		
mw	*ii*		0.5			0.5	
ii	*ii*				1		
iw	*ii*				0.5	0.5	
ww	*ii*					1	
mm	*iw*		0.5	0.5			
mi	*iw*		0.25	0.25	0.25	0.25	
mw	*iw*		0.25	0.25		0.25	0.25
ii	*iw*				0.5	0.5	
iw	*iw*				0.25	0.5	0.25
ww	*iw*					0.5	0.5
mm	*ww*			1			
mi	*ww*			0.5		0.5	
mw	*ww*			0.5			0.5
ii	*ww*					1	
iw	*ww*					0.5	0.5
ww	*ww*						1

The relative proportions of offspring genotypes from all 36 possible male x female crosses before the population is normalised. Fitness compensation is represented by the parameters G_1 and G_2, while d and k are parameters for incomplete rescue and incomplete killing respectively (taking values 0 and 1, respectively, for the simple model outlined in Section 3). Viability effects due to the costs of producing toxin and antitoxin are not shown since their effect is to reduce the numbers of all offspring of the same genotype by the same factor, e.g. all *mm* offspring amounts are scaled by a factor of $(1-Ut)^2(1-U_{at})^2$ before normalisation.

Research Paper 13. The evolutionary dynamics of male-killers and their hosts

James Randerson, Nick Smith and Laurence Hurst (1999)
Heredity, in press.

Summary

Male-killing bacteria are cytoplasmic sex ratio distorters that are transmitted vertically from female hosts to their offspring. The killing of male hosts by their bacteria is thought to be adaptive because it augments the fitness of female hosts carrying clonal relatives of those bacteria. Here we attempt to explain observations of multiple male-killers in natural host populations. First we show that such male-killer polymorphism cannot be explained by a classical model of male-killing. We then show that more complicated models incorporating the evolution of resistance in hosts can explain male-killer polymorphism. However, this is only likely if resistance genes are very costly. We also consider the long-term evolutionary dynamics of male-killers, and show that evolution towards progressively more "efficient" male-killers can be thwarted by the appearance of host resistance. The presence of a resistance gene can allow a less efficient male-killer to out-compete its rival and hence reverse the trend towards more efficient transmission and reduced metabolic load on the host.

Introduction

Male-killing bacteria belong to the class of cytoplasmic elements which spread by manipulating the sex-ratio of their hosts (Hurst, *et al.*, 1997; Hurst, 1993). Male-killers have been found in a number of bacterial genera, most notably *Wolbachia*, *Rickettsia* and *Spiroplasma* (Hurst *et al.*, 1999a; Werren, *et al.*, 1994; Williamson & Poulson, 1979), and male-killing has been reported in a wide variety of insect taxa (Hurst, et al., 1997; Hurst, 1993). In the case of "early male-killing" (*sensu* Hurst, 1991) sex-specific pathogenesis is thought to have evolved in male-killing bacteria as a consequence of their almost exclusively vertical (*i.e.* egg to brood) means of transmission between hosts (Hurst, 1991). Bacteria in male hosts are at an evolutionary dead-end, so male killing has a fitness cost of zero (from the bacterial point of view). But the death of males can augment the fitness of all remaining brood members, including female hosts carrying clonal relatives of the bacteria which killed males.

This positive effect of male-killing is termed fitness compensation, and can occur for a number of reasons: reduced intra-brood competition, reduced inbreeding or direct benefits from egg cannibalism (Hurst, 1997; and references therein). It is fitness compensation which allows male-killers to spread in a host population, despite infected male hosts being killed and infected female hosts bearing a fitness cost (Hurst, 1991).

Classical models of male-killers (Hurst, *et al.*, 1997; Hurst, 1991) have considered a number of parameters such as the vertical transmission efficiency, cost borne by infected females, and the level of fitness compensation. We consider a host population that possesses two strains of male-killers. We find that male-killer polymorphism (*sensu* Ford, 1971) is not a stable solution. As long as both male-killers are able to deterministically invade a host population free of male-killers, then only one male-killer will be maintained in the host population. The male-killer with the higher Basic Rate of Increase or *BRI* (defined later) will always out-compete the other. *BRI* is a function of the three parameters mentioned above.

This theoretical result of no male-killer polymorphism appears to be contradicted by the empirical data of a number of field studies (Hurst, *et al.*, 1999a; Majerus *et al.* Submitted). In an extreme case, four different male-killer strains were isolated from individuals of the host *Adalia bipunctata* found on a single street in Moscow (Majerus and Schulenburg *pers. com.*). How can we reconcile the theoretical and empirical data?

It is by definition, a property of all selfish elements, that their spread creates the context for the spread of host resistance genes. We ask whether the evolution of host resistance could provide conditions for male-killer polymorphism. The evolution of resistance to cytoplasmic sex-ratio distorters has received some theoretical consideration (Uyenoyama & Feldman, 1978), although not in the context of polymorphism.

Here we ask what would happen if a "weaker" male-killer were introduced into a host population which had evolved resistance to a "stronger" male-killer?

One Male Killer

We specify a model of male-killing similar to those of previous studies (Hurst, 1991; Hurst, *et al.*, 1997; Freeland & McCabe, 1997). We assume that the male-killer, MK, is transmitted to a proportion α of a female's brood. Transmission is exclusively vertical (no horizontal transmission). All infected males die, while infected females suffer a viability fitness cost *U*. Empirical evidence for such a cost has been found for *Rickettsia* infection of *Adalia Bipunctata* (Hurst, *et al.*, 1994). Fitness compensation benefits all surviving members of the brood, and is a function of the amount of male killing, which is in turn a function of α. Fitness compensation is maximised when brood fitness is unaffected by male death, so the fitness of dead males is perfectly re-distributed amongst the surviving brood. We assume fitness compensation to be a proportion ϕ of this theoretical maximum, so the compensation received by survivors as a consequence of male death is given by

$$\varphi = 1 + \phi\left(\frac{1}{1-\alpha/2} - 1\right) = 1 + \frac{\phi\alpha}{2-\alpha}. \qquad (1)$$

We assume an infinite panmictic outbred population with discrete generations. Recursion equations can be expressed in terms of infected and uninfected females since all breeding males are uninfected. The proportion of adult females infected by MK is *p*, the proportion of uninfected adult females is *q*, and *W* is the sum of the right hand sides.

$$W\,p' = p\,\alpha\,(1\text{-}U)\varphi \qquad (2)$$

$$W\,q' = p\,(1-\alpha)\varphi + q \qquad (3)$$

The invasion conditions for MK are found when:

$$\left.\frac{dp'}{dp}\right|_{p=0} > 1,$$

is satisfied, *i.e.* when the "Basic Rate of Increase" (*BRI*) is positive, where

$$BRI = \alpha\,(1\text{-}U)\,\varphi - 1. \qquad (4)$$

Note that W *does not feature in this equation because as* $p \to 0$, $W \to 1$.

This formula gives an impression of the "strength" of a male-killer, and shows that if a male-killer is to spread, then the fitness compensation ($\varphi > 1$) must be high enough to account for viability effects ($U > 0$) and imperfect transmission ($1 > \alpha > 0$).

The equilibrium value of MK (p^* is found by solving for $p' = p$, which gives

$$p^* = \frac{1 - \varphi\alpha + \varphi U\alpha}{1 - \varphi + \varphi U\alpha}. \qquad (5)$$

As reported previously (Hurst, 1991), perfect transmission of a male-killer theoretically leads to fixation of the male-killer ($p^* = 1$ when $\alpha = 1$), otherwise *p* and *q* are maintained in polymorphism ($p^* < 1$ when $\alpha < 1$). Fixation of a male-killer would lead to population extinction due to the severe shortage of males.

Two Male Killers

We now define a second male-killer, MK_2. All male-killer parameters are defined separately for $MK_1$and MK_2, except for the fitness compensation parameter ϕ (sub-scripts denote the male-killer to which each parameter applies). One could imagine a situation in which ϕ differed between male-killers, for example if killing occurred at a different time during development. However, for the sake of simplicity we assume that ϕ applies equally to both male-killers. The two male-killers are never found in the same host because there is no horizontal transmission, so the recursions are a simple extension of those for a single male-killer.

$$W\, p_1' = p_1\,(BRI_1 - 1) \qquad (6)^*$$

$$W\, p_2' = p_2\,(BRI_2 - 1) \qquad (7)$$

$$W\, q' = p_1(1-\alpha_1)\,\varphi_1 + p_2(1-\alpha_2)\,\varphi_2 + q \qquad (8)^{**}$$

The condition for MK_2 to invade a population already containing MK_1 at equilibrium is given by the solution to

$$\left.\frac{\partial p'_2}{\partial p_2}\right|_{p_1 = p_1^*} > 1,$$

which is given by

$$BRI_2 > BRI_1. \qquad (9)$$

So MK_2 can invade in the presence of MK_1 if its *BRI* is higher. Note that this result stands even if we were to assume that ϕ applies differently to the two male-killers, as ϕ is a component of each male-killer's *BRI*.

To determine whether a stable polymorphism of male-killers is possible, we need to find the intersection of $p_1'=p_1$ and $p_2'=p_2$. We find that

$$\text{at } p_1'=p_1, \qquad p_1 = p_1^* - p_2\frac{p_1^* BRI_1}{BRI_2} \qquad (10)$$

$$\text{and at } p_2'=p_2, \qquad p_1 = p_1^*\frac{BRI_2}{BRI_1} - p_2\frac{p_1^* BRI_1}{BRI_2}. \qquad (11)$$

It is clear by inspection that these lines are parallel, with gradient $p_1^*BRI_1/BRI_2$. Therefore, a neutral equilibrium is only possible if the lines have the same intercept, *i.e.* if $BRI_1=BRI_2$. Hence for stable polymorphism both male-killers must have the same *BRI*. These equations are illustrated graphically in Figure 1. Note that the direction of the vectors makes it impossible that a stable limit cycle will result.

This means that a contest between two male-killers is decided solely on which of the two has the highest *BRI*. This makes intuitive sense because in a three-allele haploid system with no

---

* *equivalent to equation (2).*

** *equivalent to equation (3) when* $p_2 = 0$.

frequency dependence, one would expect the most fit allele to simply out-compete the others. The reason that in this case the "best" male-killer does not remove the uninfected "allele" is that uninfected individuals are created each generation. The system mimics a mutation-selection equilibrium in which the uninfected "allele" is maintained by a high "mutation" rate, *i.e.* by imperfect transmission of the male-killer.

Note that it is not necessarily the case that the male-killer with the highest *BRI* will also have the highest equilibrium value, although the general trend is for *BRI* to increase with equilibrium frequency.

Resistance Genes

We considered two alternative resistance genes: a maternal-effect gene and a filial-effect gene. Both resistance genes act in the diploid host to reduce male-killing, and are inherited in an autosomal fashion. The resistance genes differ in their sex-dependent effects, with the maternal-effect gene acting to reduce vertical transmission and the filial-effect gene acting to cure males of the male-killer. Qualitatively similar results were obtained for the two models of resistance, so we present only the filial-effect resistance gene here. Details of the maternal-effect resistance can be found at URL1.

The filial-effect resistance gene (R) acts in a dominant fashion in the male host to completely eliminate male-killers from a proportion 1-g (where $0 < g < 1$) of infected male embryos, so that the proportion of males killed is $g\alpha$ rather than α. We assume that R imposes a cost c on males, and that this cost acts multiplicatively so that the fitness of males homozygous for R is $(1\text{-}c)^2$. This combination of a dominant gene with mutiplicative costs assumes that the R gene codes for an antibiotic whose effectiveness does not increase above a certain dosage, but whose production cost varies with the copy number of the R gene. The recursion equations of the model are shown in Appendix 2.

The Spread of a Resistance Gene

The conditions for R to invade a population containing a male-killer at equilibrium were obtained by modifier analysis (see URL1) and confirmed by simulation. The resistance gene can always invade if it is not costly ($c = 0$) so long as $p^* > 0$. Furthermore, R can invade even if the cost of resistance is considerable, especially if the transmission advantage of the male-killer is particularly high and if the resistance gene is highly effective (g low). The invasion conditions for a costly resistance gene are plotted in Figure 2.

Simulation shows that a costly resistance gene, if it can invade, will always reach a stable equilibrium within the population and reduce the equilibrium level of the male killer. However, this result is dependent on the population being infinitely large. If the resistance gene is highly effective (g low) then the male-killer can be reduced to such low frequencies that it would almost

certainly be lost from a finite population. We therefore incorporated into our simulations a cut-off threshold, with frequencies falling below this threshold put to zero. The results in this paper were obtained using a cut-off of 10^{-7}, but qualitatively identical results were obtained with a cut-off of 10^{-13}.

Male-killer Polymorphism With Resistance Genes

The proposed scheme for the evolution of male-killer polymorphism with resistance genes involves two steps. First, the spread of a costly resistance gene should "weaken" the resident male-killer. The results above confirm that this can happen, although sometimes both the male-killer and the resistance gene will be lost from the population. The second step is for a second male-killer, unaffected by the resistance gene, to spread at the expense of the "stronger" male-killer. The "stronger" male-killer then declines in frequency, thereby causing a decrease in the frequency of the resistance gene, and hence reducing the degree to which the "stronger" male-killer is affected by host resistance. Such frequency-dependent selection, if sufficiently damped, should allow the stable maintenance of a male-killer polymorphism. Owing to the complexity of the recursions and because of a need to incorporate cut-offs, we have used simulations to investigate whether this intuitive argument is correct.

The ability of the second male-killer to invade is not determined simply by its *BRI*. Instead, we define a General Rate of Increase (*GRI*) such that

$$GRI = \alpha (1-U) \varphi - W_f, \qquad (12)$$

where W_f is the mean female fitness, and where *GRI* has to be positive for a male-killer to invade. With no male-killers or resistance genes, $W_f=1$ and so *GRI=BRI*. However with a male-killer and resistance gene in polymorphism, W_f will be lower due to reduced male-killing (the resistance genes spread because of their effects on mean male fitness). When a male-killer (say MK_1) is at equilibrium within the population

$$W_f = \alpha_1 (1-U_1) \varphi_1. \qquad (13)$$

This means that for MK_2 to invade when MK_1 is at equilibrium, $GRI_2 > 0$ simplifies to $BRI_2 > BRI_1$, as obtained above (see Two Male Killers).

These considerations of *GRI* mean that MK_2 will always invade and eliminate MK_1 if $BRI_2 > BRI_1$, and that MK_2 is unlikely to invade unless $BRI_2>0$. This leaves an intermediate region of $BRI_1 > BRI_2 > 0$ in which male-killer polymorphism might occur. Simulations indicate that MK_2 can invade and eliminate MK_1 even if $BRI_1 > BRI_2$, and that male-killer polymorphism does occur within the bounds described. This polymorphic zone is shown in Figure 3. These figures represent the outcomes of a series of simulations in which α_1 and α_2 were varied while U_1 and U_2 were constant and equal.

To characterise parameter space more rigorously, we carried out parameter scans similar to those represented in Figure 3 for varying resistance gene and male-killer cost parameter values. An example of such a parameter scan is shown in Figure 4.

The two models of host resistance yield similar likelihoods of male-killer polymorphism. As suggested by the intuitive argument presented above, both the resistance gene's cost and its effectiveness affect the likelihood of male-killer polymorphism. As the cost of resistance c is reduced, the zone of polymorphism diminishes. A high cost is required for sufficient frequency-dependent damping. When c is high, then as resistance gene effectiveness is increased (g lowered), the likelihood of male-killer polymorphism is also increased. This is because a less effective resistance gene has a smaller effect on MK_1, thus creating a smaller potential zone of polymorphism.

Discussion

Male-killer polymorphism

We have shown that stable co-existence of two male-killer strains within a single population is impossible if no host resistance is permitted. The male-killer with the higher *BRI* will always out-compete the other and spread to its equilibrium value. In contrast, we have shown that stable male-killer polymorphism is possible if the host can evolve resistance to a male-killer. Parameter scans indicate that male-killer polymorphism is only likely if the resistance gene is costly (roughly speaking, above 1%). Such costs may seem unrealistically high, but such is the deleterious effect of the male-killer on its hosts that resistance genes many times more costly can spread.

There are alternative explanations for the observations of male-killer polymorphism. (1) Such observations may include individuals from different populations: if different male-killers are present in different populations then male-killer polymorphism may be incorrectly inferred. Proximity of individuals need not imply inter-breeding. (2) Population sub-structure might enable male-killer polymorphism, with immigration of "weaker" (*i.e.* lower *BRI*) male-killers balancing selection in favour of the "stronger" male-killers (*i.e.* higher *BRI*). (3) The observation of male-killer polymorphism does not necessarily imply a stable polymorphism: if new strains of male-killers are continuously arising then a mutation-selection balance will maintain transient polymorphisms. In the same way that we cannot assess the likelihood of the resistance gene explanation until we know about the critical parameters (such as resistance cost), these alternative models cannot be evaluated without explicit models and the necessary parameter information (such as migration and *de novo* appearance of male-killers).

What effect would one expect host resistance to have on the long-term dynamics of a host/male-killer system? Will it evolve towards a particular evolutionarily stable state? In the absence of host resistance, new male-killers can oust the incumbent male-killer if they have a higher *BRI*. This means that *BRI* has to increase over time, with an associated tendency (but not strict requirement) for male-killer equilibrium frequency to increase also. Like a ratchet mechanism, evolution can proceed stepwise in one direction, but is prevented from going in reverse. One would therefore expect male-killers in extant populations to have high transmission efficiency and impose a modest cost on their hosts.

The evolution of host resistance breaks the ratchet. If the resistance gene is costless, then the male-killer can be deterministically driven from the population. Even if the resistance gene is costly, then as long as it is highly effective then it can "cure" the population of a male-killer by reducing the male-killer down to an absorbing lower boundary.

Even if the spread of the resistance gene does not automatically eliminate the male-killer, host resistance can still cause the ratchet to click backwards. If the second male-killer is unaffected by the resistance gene, then it can eliminate the first male-killer even if it has a lower *BRI*. In many cases this second male-killer will oust the first from the population completely, which in turn will lead to the extinction of the resistance gene if costly. Host resistance can therefore cause the ratchet to click back in the opposite direction.

Empirical studies have indicated that the vertical transmission frequency of male-killers is typically in the region of 80-90% (Hurst, *et al.*, 1997; Hurst, *et al.*, 1992), although values in excess of 99% have been reported (Majerus, *et al.*, 1998). A number of ideas have been put forward to explain why transmission efficiency rarely exceeds 90% (Hurst, *et al.*, 1997).

One suggestion is that the observed transmission efficiency is the maximum the male-killer is capable of, given the constraints imposed by the host (Hurst, *et al.*, 1996). The observed transmission value may therefore be the outcome of an arms race between the bacteria and its host. A second possibility is that higher transmission efficiency is possible, but that male-killers reaching such high values of α will send their host population extinct due to the severe shortage of males. Clade selection may impose a higher order filter on evolution to very high vertical transmission efficiencies. A third suggestion is that transmission efficiency may trade off against the cost imposed on the host. This could plausibly arise if both cost and transmission are dependent on bacterial density in the host cells. From our analysis, we have shown that *BRI* depends on both U and α. Hence, if there is a trade off between α and U then one could imagine a point at which further increases in transmission efficiency actually lead to a reduction in *BRI*, and would be selectively disfavoured (Hurst, *et al.*, 1997). There is some recent evidence for a link between bacterial load and CI level for *Wolbachia* in *Nasonia vitripennis* (*e.g.* Perrot-Minnot & Werren, 1999).

Our results suggest three further possibilities. Firstly, in the absence of resistance, we show that one male-killer can eliminate another male-killer by having a higher *BRI*, despite having a lower transmission efficiency. Secondly, the spread of host resistance can lead to complete elimination of a male-killer. Thirdly, we show that if host resistance has evolved to a resident male-killer, a second male-killer can eliminate the resident male-killer even with a lower *BRI* and lower transmission efficiency.

The evolution of host resistance

We have shown that the evolution of host resistance affects both the likelihood of the male-killer polymorphism and the long-term dynamics of male-killers and their hosts. To what extent are our results dependent on our specific models of host resistance, and is their any evidence of resistance genes in nature?

Although conclusive evidence of host resistance to male-killers is lacking, a number of studies have suggested that the effectiveness of a male-killer varies with host genotype (Cavalcanti, *et al.*, 1957; Malagolowkin & Poulson, 1957).

A huge variety of alternative resistance genes are imaginable. We have chosen two models of resistance, which we consider to represent the middle ground, in that it is possible to imagine both resistance genes that will spread more easily, and resistance genes that will spread less easily.

Why might the evolution of host resistance be harder than suggested by our models? Hurst *et al.* (1997) consider the consequences of a resistance gene causing increased longevity in males. If resistance gene increases male life span by only a small time, it may actually be selectively disfavoured. This is because a male that dies as a larva is still incapable of passing on his genes, and furthermore contributes less fitness compensation to the surviving members of the brood. In order to spread, a resistance gene must allow some males to reach maturity. We have neglected this problem, since the fitness compensation parameter is unaffected by the resistance genes. In effect, we consider that our resistance genes enable a complete transition from infected to uninfected, without any intermediate classes.

On the other hand, neither of our resistance genes affect the cost of male-killers on females (U). If a resistance gene could reduce U as well as the transmission of male-killers, then its spread would be more strongly favoured.

There is a pressing need for empirical data to back up the theoretical approach adopted in this paper. We need to know if resistance genes exist, and if so, we need to know how the resistance genes act. It seems reasonable to suppose that the qualitative results reached in this paper (male-killer polymorphism with resistance genes is possible) are probably independent of the exact form of a resistance gene. Any quantitative results however (resistance gene cost must be of the order of 10^{-2} or above), are highly dependent on the exact nature of host resistance to male-killers.

Acknowledgements

The authors wish to thank Greg Hurst, Mike Majerus, Hinrich Von Der Schulenburg and Tamsin Majerus for helpful comments on the manuscript and general advice. Also the authors would like to thank two anonymous reviews and the corresponding editor for helpful comments on the manuscript. LDH is funded by a Royal Society fellowship and JPR is funded by a NERC studentship.

Appendix 1 - List of Parameters and abbreviations.

MK_1- male-killer one.

MK_2- male-killer two.

p_1- frequency of MK_1.

p_2- frequency of MK_2.

q- frequency of uninfected individuals.

α- transmission efficiency of the MK.

U- cost imposed on infected females.

ϕ- proportion of the theoretical maximum fitness compensation received by surviving brood members.

φ- fitness compensation to surviving brood due to male death.

p^*- equilibrium frequency of a particular bacterial strain in a non-resistant host population.

BRI- Basic Rate of Increase of a particular male killer.

R- filial-effect resistance gene.

c- cost imposed by a resistance gene on the sex in which it acts.

g – reduction in bacterial action in infected males with the R gene.

d – reduction in bacterial transmission from mother to eggs due to maternal resistance (see URL1).

Appendix 2 – Details of the Filial-Resistance Model

The genotype frequencies are represented as follows:

males: y, females: x

	No MK	MK_1 (females only)	MK_2 (females only)
No R gene	y_1, x_1	x_4	x_7
Heterozygous for R gene	y_2, x_2	x_5	x_8
Homozygous for R gene	y_3, x_3	x_6	x_9

Recursion equations for the dynamics of a two male-killer system with the filial-effect resistance gene:

	y_1	y_2	y_3
1	$y_1 + x_1$	$\frac{1}{2}(y_1 + y_2 + x_1 + x_2)$	$y_2 + x_2$
2	$\frac{1}{2}(y_1 + y_2 + x_1 + x_2)$	$y_1/4 + y_2/2 + y_3/4 + x_1/4 + x_2/2 + x_3/4$	$\frac{1}{2}(y_2 + y_3 + x_2 + x_3)$
3	$y_2 + x_2$	$\frac{1}{2}(y_2 + y_3 + x_2 + x_3)$	$y_3 + x_3$
4	$y_1(1-\alpha_1) + x_4\alpha_1 + x_1(1-\alpha_1)$	$\frac{1}{2}(y_1(1-\alpha_1) + y_2(1-g\alpha_1) + x_4\alpha_1 + x_5\alpha_1 + x_1(1-\alpha_1) + x_2(1-\alpha_1))$	$y_2(1-g\alpha_1) + x_5\alpha_1 + x_2(1-\alpha_1)$
5	$\frac{1}{2}(y_1(1-\alpha_1) + y_2(1-g\alpha_1) + x_4\alpha_1 + x_5\alpha_1 + x_1(1-\alpha_1) + x_2(1-\alpha_1))$	$\frac{1}{4}(y_1(1-\alpha_1) + 2y_2(1-g\alpha_1) + y_3(1-g\alpha_1) + x_4\alpha_1 + 2x_5\alpha_1 + x_6\alpha_1 + x_1(1-\alpha_1) + 2x_2(1-\alpha_1) + x_3(1-\alpha_1))$	$\frac{1}{2}(y_2(1-g\alpha_1) + y_3(1-g\alpha_1) + x_5\alpha_1 + x_6\alpha_1 + x_2(1-\alpha_1) + x_3(1-\alpha_1))$
6	$y_2(1-g\alpha_1) + x_5\alpha_1 + x_2(1-\alpha_1)$	$\frac{1}{2}(y_2(1-g\alpha_1) + y_3(1-g\alpha_1) + x_5\alpha_1 + x_6\alpha_1 + x_2(1-\alpha_1) + x_3(1-\alpha_1))$	$y_3(1-g\alpha_1) + x_6\alpha_1 + x_3(1-\alpha_1)$
7	$y_1(1-\alpha_2) + x_7\alpha_2 + x_1(1-\alpha_2)$	$\frac{1}{2}(y_1(1-\alpha_2) + y_2(1-\alpha_2) + x_7\alpha_2 + x_8\alpha_2 + x_1(1-\alpha_2) + x_2(1-\alpha_2))$	$y_2(1-\alpha_2) + x_8\alpha_2 + x_2(1-\alpha_2)$
8	$\frac{1}{2}(y_1(1-\alpha_2) + y_2(1-\alpha_2) + x_7\alpha_2 + x_8\alpha_2 + x_1(1-\alpha_2) + x_2(1-\alpha_2))$	$\frac{1}{4}(y_1(1-\alpha_2) + 2y_2(1-\alpha_2) + y_3(1-\alpha_2) + x_7\alpha_2 + 2x_8\alpha_2 + x_9\alpha_2 + x_1(1-\alpha_2) + 2x_2(1-\alpha_2) + x_3(1-\alpha_2))$	$\frac{1}{2}(y_2(1-\alpha_2) + y_3(1-\alpha_2) + x_8\alpha_2 + x_9\alpha_2 + x_2(1-\alpha_2) + x_3(1-\alpha_2))$
9	$y_2(1-\alpha_2) + x_8\alpha_2 + x_2(1-\alpha_2)$	$\frac{1}{2}(y_2(1-\alpha_2) + y_3(1-\alpha_2) + x_8\alpha_2 + x_9\alpha_2 + x_2(1-\alpha_2) + x_3(1-\alpha_2)$	$y_3(1-\alpha_2) + x_9\alpha_2 + x_3(1-\alpha_2)$

Male death occurs in broods from mothers $x_4 - x_9$, and the number of males that dying (and hence the amount of fitness compensation) is dependent both on the mother and the father genotype.

Parental Genotypes	Appropriate fitness compensation term
$y_1 x_4$	$1+\frac{\phi\alpha_1}{2-\alpha_1}$
$y_1 x_6$, $y_2 x_6$, $y_3 x_4$, $y_3 x_5$, $y_3 x_6$	$1+\frac{\phi g\alpha_1}{2-g\alpha_1}$
$y_2 x_5$	$1+\frac{(0.75g\alpha_1+0.25\alpha_1)\phi}{2-(0.75g\alpha_1+0.25\alpha_1)}$
$y_1 x_5$, $y_2 x_4$,	$1+\frac{(0.5g\alpha_1+0.5\alpha_1)\phi}{2-(0.5g\alpha_1+0.5\alpha_1)}$
All matings involving x_7, x_8 and x_9	$1+\frac{\phi\alpha_2}{2-\alpha_2}$

Females in the next generation then suffer viability costs dependent on their genotype. In females it is the cost imposed by the male-killer, while males suffer the viability cost of R.

Genotype	Viability costs
$x_1 x_2 x_3 y_1$	-
y_2	$(1-c)$
y_3	$(1-c)^2$
$x_4 x_5 x_6$	$(1-U_1)$
$x_7 x_8 x_9$	$(1-U_2)$

The invasion conditions for the R gene may be obtained by modifier analysis. The resulting expression is too long to reproduce here, but can be viewed at URL1. However, see Figure 2 for a graphical representation.

References

CAVALCANTI, A. L. G., N, F. D. AND CASTRO, L. E. 1957. Sex ratio in *Drosophila prosaltans* a character due to interaction between nuclear genes and cytoplasmic factors. *Am. Nat.*, **91**, 327-329.

FORD, E. B. 1971. *Ecological Genetics*, Chapman and Hall, London.

FREELAND, S. J. AND McCABE, B. K. 1997. Fitness compensation and the evolution of selfish cytoplasmic elements. *Heredity*, **78**, 391-402.

HURST, L. D. 1991. The incidences and evolution of cytoplasmic male killers. *Proc. R. Soc. Lond. B*, **244**, 91-99.

HURST, L. D. 1993. The incidences, mechanisms and evolution of cytoplasmic sex ratio distorters in animals. *Biological Reviews*, **68**, 121-193.

HURST, G. D. D., MAJERUS, M. E. N. AND WALKER, L. E. 1992. Cytoplasmic male killing elements in *Adalia bipunctata* (linnaeus) (coleoptera, coccinellidae). *Heredity*, **69**, 84-91.

HURST, G. D. D., PURVIS, E. L., SLOGGETT, J. J. AND MAJERUS, M. E. N. 1994. The effect of infection with male-killing *rickettsia* on the demography of female *Adalia bipunctata* l. (2-spot ladybird). *Heredity*, **73**, 309-316.

HURST, G. D. D., HAMMARTON, T. C., OBRYCKI, J. J., MAJERUS, T. M. O., WALKER, L. E., BERTRAND, D. AND MAJERUS, M. E. N. 1996. Male-killing bacterium in a 5th ladybird beetle, *Coleomegilla maculata* (coleoptera, coccinellidae). *Heredity*, **77**, 177-185.

HURST, G. D. D., HURST, L. D. AND MAJERUS, M. E. N. 1997. Cytoplasmic sex ratio distorters. In: Hoffmann, A., O'Neill, S. and Werren, J. (Eds) *Influential Passengers*, pp. 125-154. Oxford University Press, Oxford.

HURST, G. D. D., JIGGINS, F. M., SCHULENBURG, J. H. G. V. D., BERTRAND, D., WEST, S. A., GORIACHEVA, I. I., ZAKHAROV, I. A., WERREN, J. H., STOUTHAMER, R. AND MAJERUS, M. E. N. 1999a. Male-killing *Wolbachia* in two species of insect. *Proc. R. Soc. Lond. B*, **266**, 735-740.

HURST, G. D. D., SCHULENBURG, J. H. G. V. D., MAJERUS, T. M. O., BERTRAND, D., ZAKHAROV, I. A., BAUNGAARD, J., VOLKL, W., STOUTHAMER, R. AND MAJERUS, M. E. N. 1999b. Invasion of one insect species, *Adalia bipunctata*, by two different male-killing bacteria. *Insect Molecular Biology*, **8**, 133-139.

MAJERUS, M. E. N, SCHULENBURG, J. H. G. V. D., AND ZAKHAROV, I. A. Submitted. Multiple cause of male-killing in a single sample of the 2 spot ladybird, *Adalia bipunctata* (Coleoptera: Coccinellidae) from Moscow.

MAJERUS, T. M. O., MAJERUS, M. E. N., KNOWLES, B., WHEELER, J., BERTRAND, D., KUZNETZOV, V. N., UENO, H. AND HURST, G. D. D. 1998. Extreme variation in the prevalence of inherited male-killing microorganisms between three populations of *Harmonia axyridis* (Coleoptera: Coccinellidae). *Heredity*, **81**, 683-691.

MALAGOLOWKIN, C. AND POULSON, D. F. 1957. Infective transfer of maternally inherited abnormal "Sex ratio" in *Drosophila willistoni*. *Science*, **126**, 32.

PERROT-MINNOT, M.-J. AND WERREN, J. H. 1999. *Wolbachia* infection and incompatibility dynamics in experimental selection lines. *Journal of Evolutionary Biology*, **12**, 272-282.

UYENOYAMA, M. K. AND FELDMAN, M. W. 1978. The genetics of sex ratio distortion by cytoplasmic infection under maternal and contagious transmission: An epidemiological study. *Theor. Pop. Biol.*, **14**, 471-497.

WERREN, J. H., HURST, G. D. D., ZHANG, W., BREEUWER, J. A. J., STOUTHAMER, R. AND MAJERUS, M. E. N. 1994. Rickettsial relative associated with male killing in the ladybird beetle (*Adalia bipunctata*). *Journal of Bacteriology*, **176**, 388-394.

WERREN, J. H., ZHANG, W. AND GUO, L. R. 1995. Evolution and phylogeny of *Wolbachia* - reproductive parasites of arthropods. *Proc R Soc Lond B*, **261**, 55-63.

WILLIAMSON, D. L. AND POULSON, D. F. 1979. Sex ratio organisms (Spiroplasmas) of *Drosophila*. In: Whitcomb, R. F. and Tully, J. G. (Eds) *The Mycoplasmas*, pp. 175-208. Academic Press, New York.

URL1 - http://www.bath.ac.uk/Departments/BiolBioch/hurst.htm

Figure Legends

Figure 1. Illustration that stable male-killer polymorphism is impossible in a panmictic, non-resistant population. The graph shows plots of p_1 (the frequency of MK_1) against p_2 (the frequency of MK_2) for $p_1' = p_1$ and $p_2' = p_2$. The only stable points for the system are at $p_1 = 0$ and $p_2 = p_2^*$ or $p_1 = p_1^*$ and $p_2 = 0$, unless $BRI_1 = BRI_2$. So, a neutral equilibrium is only possible if the *BRI* of the two MK's is the same. However, this doesn't necessarily mean that α and U must be the same.

Figure 2. Invasion conditions for the resistance gene in the presence of a male-killer at equilibrium ($\phi = 0.5$ and $U_I = 0.01$). For invasion, the cost of the resistance gene must lie beneath the sheet. Note that as the transmission efficiency of MK_1 increases, the invasion conditions become less stringent. This is because transmission of the gene through saved males becomes increasingly significant as the population becomes more female biased. The transmission reduction parameter g has little effect on the cost tolerated by the invasion conditions except when g approaches unity. Invasion cannot occur at $g = 1$ because this represents zero transmission reduction.

Figure 3. Plot illustrating the parameter space in which we observe male-killer polymorphism. It is a representation of a series of simulation runs for both resistance genes (filial-effect and maternal-effect) at different combinations of α_1 and α_2. For all simulations $\phi = 0.5$ and $U_I = U_2 = 0.01$. Also, the characteristics of the two resistance genes are held constant ($g = d = 0.5$, $c = 0.01$). For the parameter values specified, a male-killer cannot invade if $\alpha < 0.769$ because its *BRI* will be less than zero. Polymorphism is therefore impossible if MK_2 has a transmission efficiency of less than 0.769. Polymorphism is also impossible above the line $\alpha_1 = \alpha_2$ because if MK_2 has the higher *BRI* then it will always oust MK_1 from the population regardless of host resistance. The presence of the resistance gene allows MK_2 to completely out-compete MK_1 in some circumstances, even if it has a lower *BRI*. However, near the lower boundary of the region in which MK_2 can invade, it is unable to oust MK_1 from the population. This results in stable maintenance of the two male-killers. The region enclosed by the open dotted lines represents parameter space in which polymorphism results for the filial-effect model (that described in the text). The equivalent region for the maternal-effect model is enclosed by the closed dotted lines (see URL1).

Figure 4. Out-come of a more extensive scan of parameter space. Each point on the graph represents the area of the polymorphic zone in individual scans such as Figure 3. These individual scans were undertaken for quantitatively different resistance genes so g, d and c were varied. For each scan there are 2784 different combinations of parameter values that could theoretically allow polymorphism. These lie between the lowest α value that will allow invasion and the line $\alpha_1 = \alpha_2$.

The parameter space area on the y-axis is therefore the number of these parameter combinations that resulted in polymorphism. The closed symbols represent results for the maternal-effect gene and the open symbols represent results for the filial-effect gene. The different shaped symbols represent different values of *g* and *d*: circle = 0.2, triangle = 0.5 and square = 0.8. The major trends visible in the scan are that polymorphism is less likely for a low cost resistance gene, and polymorphism is more likely for a more effective resistance gene (*i.e.* one which reduces transmission or action by a lot). It is evident that neither class of resistance gene is substantially more likely to result in polymorphism.

Figure 1

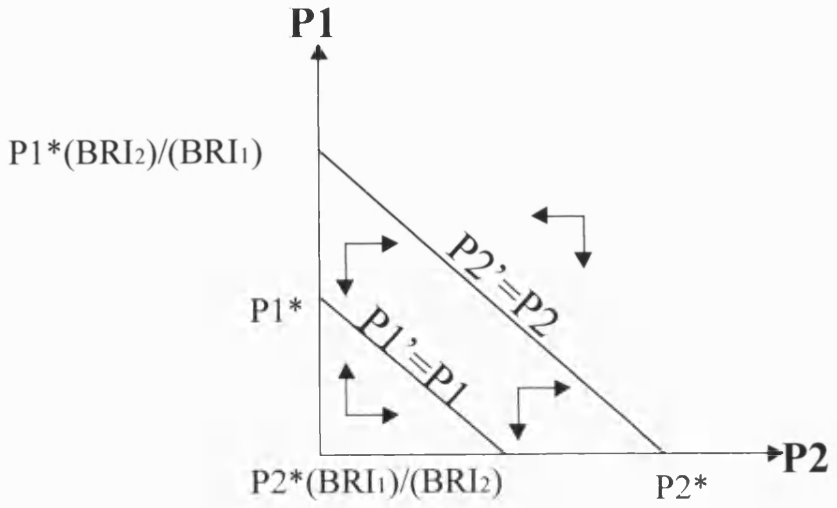


Figure 2

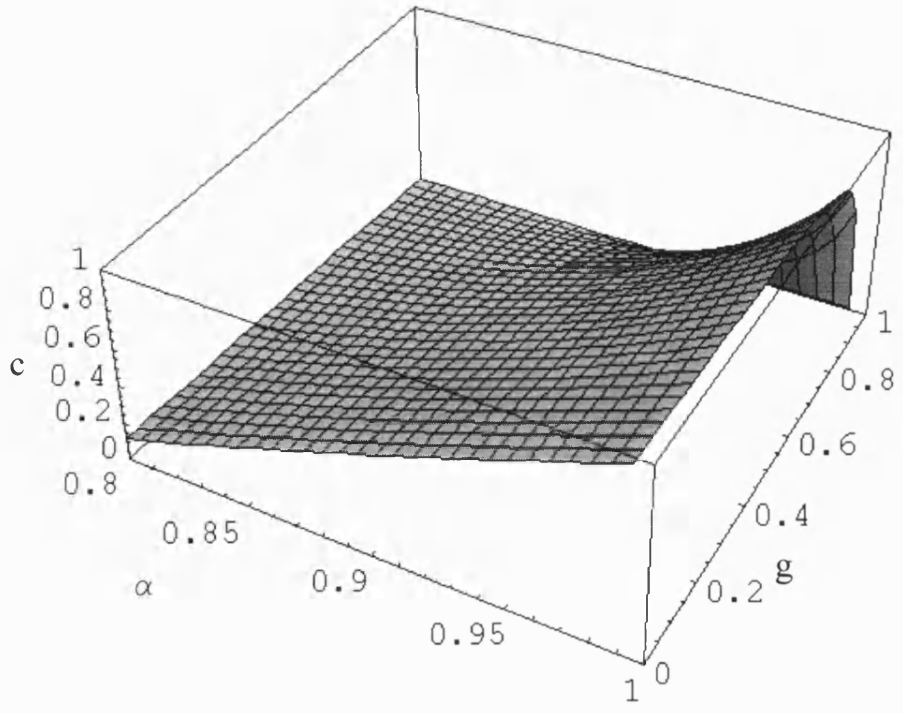


Figure 3

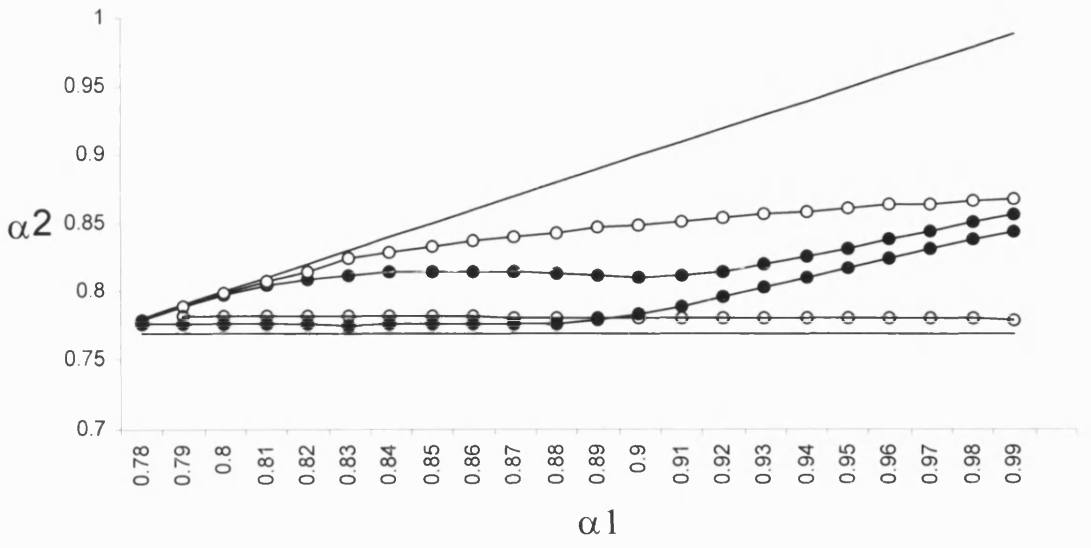
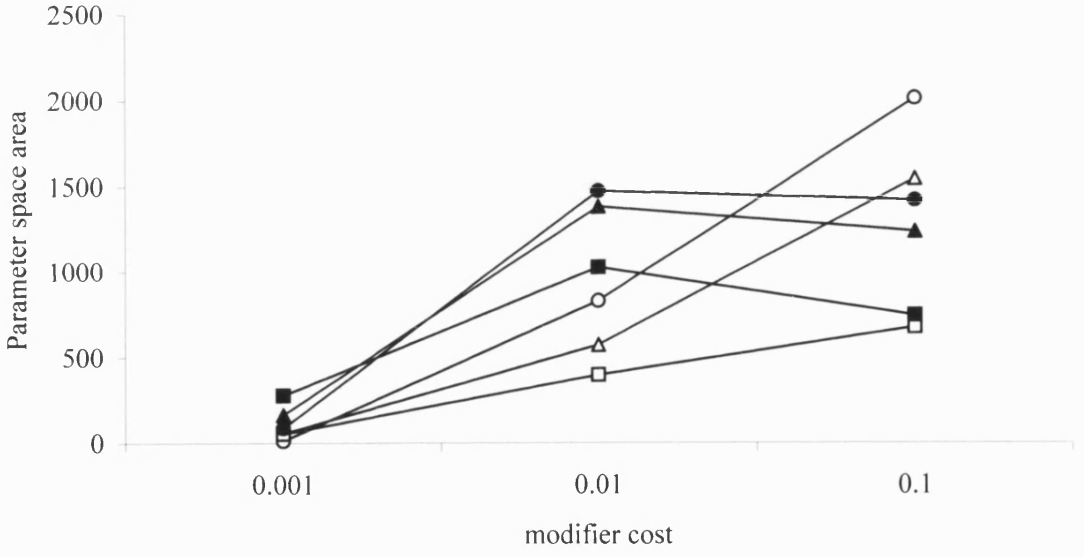


Figure 4



References for Thesis Chapters

Adzhubei, A. A., I. A. Adzhubei, I. A. Krasheninnikov and S. Neidle, 1996 Non-random usage of 'degenerate' codons is related to protein three-dimensional structure. FEBS Letters **399:** 78-82.

Akashi, H., 1994 Synonymous codon usage in *Drosophila melanogaster* - natural selection and translational accuracy. Genetics **136:** 927-935.

Akashi, H., 1995 Inferring weak selection from patterns of polymorphism and divergence at silent sites in *Drosophila* DNA. Genetics **139:** 1067-1076.

Akashi, H., 1999 Inferring the fitness effects of DNA mutations from polymorphism and divergence data: statistical power to detect directional selection under stationarity and free recombination. Genetics **151:** 221-238.

Akashi, H., and A. Eyre-Walker, 1998 Translational selection and molecular evolution. Curr. Opin. Genet. Dev. **8:** 688-693.

Alakokko, L., A. P. Kvist, M. Metsaranta, K. I. Kivirikko, B. Decrombrugghe *et al.*, 1995 Conservation of the sizes of 53 introns and over 100 intronic sequences for the binding of common transcription factors in the human and mouse genes for type II procollagen (Col2a1). Biochemical Journal **308:** 923-929.

Altschul, S. F., 1997 Sequence comparison and alignment., pp. 137-168 in *DNA and Protein sequence analysis*, edited by M. J. Bishop, and C. J. Rawlings. Oxford University Press, Oxford.

Alvarez-Valin, F., K. Jabbari and G. Bernardi, 1998 Synonymous and nonsynonymous substitutions in mammalian genes: intragenic correlations. J. Mol. Evol. **46:** 37-44.

Amundson, R., 1996 Historical development of the concept of adaptation, pp. 11-54 in *Adaptation*, edited by M. R. Rose, and G. V. Lauder. Academic Press, San Diego, California, USA.

Aota, S., and T. Ikemura, 1986 Diversity in G&C content at the 3rd position of codons in vertebrate genes and its cause. Nucleic Acid Res. **14:** 6345-6355.

Arendt, J. D., 1997 Adaptive intrinsic growth rates: an integration across taxa. Q. Rev. Biol. **72:** 149-177.

Arnheim, N., 1983 Concerted evolution of multigene families, pp. 38-61 in *Evolution of Genes and Proteins*, edited by M. Nei, and R. K. Koehn. Sinauer Associates, Sunderland, MA.

Ayala, F. J., 1999 Molecular clock mirages. Bioessays **21:** 71-75.

Bailey, W. J., J. Kim, G. P. Wagner and F. H. Ruddle, 1997 Phylogenetic reconstruction of vertebrate Hox cluster duplications. Mol. Biol. Evol. **14:** 843-853.

Barlow, D. P., 1993 Methylation and imprinting: from host defense to gene regulation? Science **260:** 309-310.

Barrow, J. D., and F. J. Tipler, 1996 *The Antropic Cosmological Principle.* Oxford University Press, Oxford.

Bartolomei, M. S., 1997 Function and epigenetic modification of the imprinted H19 gene, pp. in *Genomic imprinting*, edited by W. Reik, and A. Surani. Oxford University Press, Oxford.

Barton, N. H., 1995 Linkage and the limits to natural selection. Genetics **140:** 821-841.

Beeman, R. W., K. S. Friesen and R. E. Denell, 1992 Maternal-effect selfish genes in flour beetles. Science **256:** 89-92.

Bell, G., 1982 *The Masterpiece of Nature: The Evolution and Genetics of Sexuality.* University of California Press, Berkeley.

Bell, G., 1997 The evolution of the life cycle of brown seaweeds. Biol. J. Linn. Soc. **60:** 21-38.

Bengtsson, B., 1992 Deleterious mutations and the origin of the meiotic ploidy cycle. Genetics **131:** 741-744.

Berg, D. E., and M. M. Howe (Editors), 1989 *Mobile DNA.* American Society for Microbiology, Washington, DC.

Berg, O. G., and C. G. Kurland, 1997 Growth rate optimised tRNA abundance and codon usage. J. Mol. Biol. **270:** 544-550.

Bernardi, G., 1995 The human genome: Organization and evolutionary history. Annu. Rev. Genet. **29:** 445-476.

Bernardi, G., D. Mouchiroud and C. Gautier, 1997 Isochores and synonymous substitutions in mammalian genes., pp. 137-168 in *DNA and Protein sequence analysis*, edited by M. J. Bishop, and C. J. Rawlings. Oxford University Press, Oxford.

Bestor, T. H., 1998 Cytosine methylation and the unequal developmental potentials of the oocyte and sperm genomes. Am. J. Hum. Genet. **62:** 1269-1273.

Bird, A. P., 1995 Gene number, noise reduction and biological complexity. Trends Genet. **11:** 94-99.

Blumenthal, T., 1995 Trans-splicing and polycistronic transcription in Caenorhabditis elegans. Trends Genet. **11:** 132-136.

Blumenthal, T., 1998 Gene clusters and polycistronic transcription in eukaryotes. Bioessays **20:** 480-487.

Bodmer, W. F., and P. A. Parsons, 1962 Linkage and recombination in evolution. Adv. Genet. **11:** 1-100.

Borges, J. L., 1941 The Library of Babel, *Collected Fictions.* Penguin Books, London.

Borgia, G., 1980 Evolution of haplodiploidy: models for inbred and outbred systems. Theor. Popul. Biol. **17:** 103-128.

Borst, P., and I. Chaves, 1999 Mono-allelic expression of genes in simple eukaryotes. Trends Genet. **15:** 95-96.

Bourguet, D., and M. Raymond, 1998 The molecular basis of dominance relationships: the case of some recent adaptive genes. J. Evol. Biol. **11:** 103-122.

Brannan, C. I., and M. S. Bartolomei, 1999 Mechanisms of genomic imprinting. Current Opinion in Genetics & Development **9:** 164-170.

Britten, R. J., 1993 Forbidden synonymous substitutions in coding regions. Mol. Biol. Evol. **10:** 205-220.

Britten, R. J., and E. H. Davidson, 1971 Repetitive and non-repetitive sequences and the origin of evolutionary novelty. Q. Rev. Biol. **46:** 111-113.

Brodsky, V. Y., and I. V. Uryvaeva, 1985 *Genome multiplication in growth and development.* Cambridge University Press, Cambridge.

Brogna, S., and M. Ashburner, 1997 The Adh-related gene of Drosophila melanogaster is expressed as a functional dicistronic messenger RNA: multigenic transcription in higher organisms. EMBO J **16:** 2023-2031.

Brookfield, J. F. Y., 1997a Genetic redundancy: screening for selection in yeast. Current Biology **7:** R366-R368.

Brookfield, J. F. Y., 1997b Hypothesis testing in evolutionary inference. J.Theor.Biol. **185:** 533-538.

Brown, S. W., 1964 Automatic frequency response in the evolution of male haploidy and other coccid chromosomal systems. Genetics **49:** 797-817.

Brown, S. W., and U. Nur, 1964 Heterochromatic chromosmes in the coccids. Science 130-137.

Bull, J. J., 1979 An advantage for the evolution of male haploidy and systems with similar genetic transmission. Heredity **43:** 361-381.

Bull, J. J., 1983 *The evolution of sex determining mechanisms.* Benjamin Cummings, Menlo Park, California.

Bulmer, M., 1986 Neighboring base effects on substitution rates in pseudogenes. Mol. Biol. Evol. **3:** 322-329.

Bulmer, M., 1990 The effects of context on synonymous codon usage in genes with low codon usage bias. Nucleic Acid Res. **18:** 2869-2873.

Bulmer, M., 1991 The selection-mutation-drift theory of synonymous codon usage. Genetics **129:** 897-907.

Burt, A., and R. Trivers, 1998 Genetic conflicts in genomic imprinting. Proc. R. Soc. Lond. B **265:** 2393-2397.

Carpousis, A. J., N. F. Vanzo and L. C. Raynal, 1999 mRNA degradation. Trends Genet. **15:** 24-28.

Casane, D., S. Boissinot, B. H. J. Chang, L. C. Shimmin and W. H. Li, 1997 Mutation pattern variation among regions of the primate genome. J. Mol. Evol. **45:** 216-226.

Cavalier-Smith, T., 1978 Nuclear volume control by nucleoskeletal DNA, selection for cell volume and cell growth rate, and the solution of the DNA C-value paradox. J. Cell Sci. **34:** 247-278.

Cavalier-Smith, T., 1985 *The Evolution of Genome Size.* John Wiley, Chichester.

Chan, L., 1993 RNA editing: exploring one mode with apolipoprotein B mRNA. Bioessays **15:** 33-41.

Chao, L., and E. C. Cox, 1983 Competition between high and low mutating strains of *Escherichia coli*. Evolution **37:** 125-134.

Charlesworth, B., 1978 Model for the evolution of Y chromosomes and dosage compensation. Proc. Natl. Acad. Sci. USA **75:** 5618-5622.

Charlesworth, B., 1994 Patterns in the genome. Current Biology **4:** 182-184.

Charlesworth, B., and D. Charlesworth, 1998 Some evolutionary consequences of deleterious mutations. Genetica **103:** 3-19.

Charlesworth, B., J. A. Coyne and N. H. Barton, 1987 The relative rates of evolution of sex chromosomes and autosomes. Am. Nat. **130:** 113-146.

Charlesworth, B., M. T. Morgan and D. Charlesworth, 1993 The effect of deleterious mutations on neutral molecular variation. Genetics **134:** 1289-1303.

Charlesworth, B., P. Sniegowski and W. Stephan, 1994 The evolutionary dynamics of repetitive DNA in eukaryotes. Nature **371:** 215-220.

Charlesworth, D., and B. Charlesworth, 1978 A model for the evolution of dioecy and gynodioecy. Am Nat **112:** 975-997.

Charnov, E., 1982 *The Theory of Sex Allocation.* Princeton University Press, Princeton.

Clark, A. G., 1994 Invasion and maintenance of a gene duplication. Proc. Natl. Acad. Sci. USA **91:** 2950-2954.

Clark, M. S., 1999 Comparative genomics: the key to understanding the Human Genome Project. Bioessays **21:** 121-130.

Comeron, J. M., and M. Aguade, 1996 Synonymous substitutions in the Xdh gene of *Drosophila* - heterogeneous distribution along the coding region. Genetics **144:** 1053-1062.

Comeron, J. M., and M. Aguade, 1998 An evaluation of measures of synonymous codon usage bias. J. Mol. Evol. **47:** 268-274.

Comeron, J. M., M. Kreitman and M. Agaude, 1999 Natural slection on synonymous sites is correlated with gene length and recombination in Drosophila. Genetics **151:** 239-249.

Conn, G. L., and D. E. Draper, 1998 RNA structure. Curr. Opin. Struct. Biol. **8:** 278-285.

Constancia, M., B. Pickard, G. Kelsey and W. Reik, 1998 Imprinting mechanisms. Gen. Res. **8:** 881-900.

Cooke, J., M. A. Nowak, M. Boerlijst and J. Maynard-Smith, 1997 Evolutionary origins and maintenance of redundant gene expression during metazoan development. Trends Genet. **13:** 360-364.

Cosmides, L. M., and J. Tooby, 1981 Cytoplasmic inheritance and intragenomic conflict. J. theor. Biol. **89:** 83-129.

Crow, J., and M. Kimura, 1970 *An Introduction to Population Genetics Theory.* Harper and Rowe, New York.

Crow, J. F., 1958 Some possibilities for measuring selection intensities in man. Human Biology **30:** 1-13.

Crow, J. F., 1991 Why Is Mendelian segregation so exact? Bioessays **13:** 305-312.

Crow, J. F., and M. Kimura, 1965 Evolution in sexual and asexual populations. Am. Nat **99:** 439-450.

Crozier, R. H., and P. Pamilo, 1996 *Evolution of Social Insect Colonies: Sex Allocation and Kin Selection.* Oxford University Press, Oxford.

Culbertson, M. R., 1999 RNA surveillance - unforeseen consequences for gene expression, inherited genetic disorders and cancer. Trends Genet. **15:** 74-80.

Dandekar, T., B. Snel, M. Huynen and P. Bork, 1998 Conservation of gene order: a fingerprint of proteins that physically interact. Trends. Biochem. Sci. **23:** 324-328.

David, P., 1997 Modeling the genetic basis of heterosis: tests of alternative hypotheses. Evolution **51:** 1049-1057.

Dawkins, R., 1976 *The Selfish Gene.* Oxford University Press, Oxford.

Dawkins, R., 1990 Parasites, desiderata lists and the paradox of the organism. Parasitology **100:** S63-S73.

Dawson, K. J., 1998 Evolutionarily stable mutation rates. J. Theor Biol **194:** 143-157.

Debry, R. W., and W. F. Marzluff, 1994 Selection on silent sites in the rodent H3 histone gene family. Genetics **138:** 191-202.

Demerec, M., and P. Hartman, 1959 Complex loci in microorganisms. Annu. Rev. Microbiol. **13:** 377-406.

Deng, H. W., Y. X. Fu and M. Lynch, 1998 Inferring the major genomic mode of dominance and overdominance. Genetica **103:** 559-567.

Dennet, D. C., 1995 *Darwin's Dangerous Idea.* Penguin Books, London.

Descartes, R., 1637 *Discourse on the Method and the Meditations.* Penguin Books, London.

Dimitri, P., and N. Junakovic, 1999 Revising the selfish DNA hypothesis - new evidence on accumulation of transposable elements in heterochromatin. Trends Genet. **15:** 123-124.

Donofrio, G., D. Mouchiroud, B. Aissani, C. Gautier and G. Bernardi, 1991 Correlations between the compositional properties of human genes, codon usage, and amino acid composition of proteins. J. Mol. Evol. **32:** 504-510.

Doolittle, W. F., and C. Sapienza, 1980 Selfish genes, the phenotype paradigm and genome evolution. Nature **284:** 601-603.

Douglas, A. E., 1989 Mycetocyte symbiosis in insects. Biol. Rev. **64:** 409-434.

Dover, G., 1982 Molecular drive: a cohesive mode a species evolution. Nature **299:** 111-116.

Drake, J. W., 1991 A constant rate of spontaneous mutation in DNA-based microbes. Proc. Natl. Acad. Sci. USA **88:** 7160-7164.

Duret, L., D. Mouchiroud and C. Gautier, 1995 Statistical analysis of vertebrate sequences reveals that long genes are scarce in GC rich isochores. J. Mol. Evol. **40:** 308-317.

Duret, L., D. Mouchiroud and M. Gouy, 1994 Hovergen - a database of homologous vertebrate genes. Nucleic Acid Res. **22:** 2360-2365.

Efstratiadis, A., 1994 Parental imprinting of autosomal mammalian genes. Curr. Opin. Genet. Devel. **4:** 265-280.

Elder, J. F., and B. J. Turner, 1996 Concerted evolution of repetitive DNA sequences in eukaryotes. Q. Rev. Biol. **71:** 79.

Elena, S. F., L. Ekunwe, N. Hajela, S. A. Oden and R. E. Lenski, 1998 Distribution of fitness effects caused by random insertion mutations in Escherichia coli. Genetica **103:** 349-358.

Elena, S. F., and R. E. Lenski, 1997 Test of synergistic interactions among deleterious mutations in bacteria. Nature **390:** 395-397.

Elgin, S. C. R., and H. Weintraub, 1975 Chromosomal proteins and chromatin structure. Ann. Rev. Biochem. **44:** 725-774.

Ellis, N. A., 1998 The war of the sex chromosomes. Nature Genet. **20:** 9-10.

Endo, T., T. Imanishi, T. Gojobori and H. Inoko, 1997 Evolutionary significance of intra-genome duplications on human chromosomes. Gene **205:** 19-27.

Eyre-Walker, A., 1991 An analysis of codon usage in mammals: selection or mutation bias? J. Mol. Evol. **33:** 442-449.

Eyre-Walker, A., 1993 Recombination and mammalian genome evolution. Proc. Roy. Soc. Lond. B. **252:** 237-243.

Eyre-Walker, A., 1996 Synonymous codon bias is related to gene length in *Escherichia coli*: selection for translational accuracy? Mol. Biol. Evol. **13:** 864-872.

Eyre-Walker, A., and M. Bulmer, 1993 Reduced synonymous substitution rate at the start of enterobacterial genes. Nucleic Acid Res. **21:** 4599-4603.

Eyre-Walker, A., and P. D. Keightley, 1999 High genomic deleterious mutation rates in hominids. Nature **397:** 344-347.

Fan, J.-B., S. H. Korman, C. R. Cantor and C. L. Smith, 1991 *Giardia lamblia*: haploid genome size determined by pulsed field gel electrophoresis is less than 12 Mb. Nucleic Acids Res. **19:** 1905-1908.

Feynman, R., 1966 Nobel prize lecture. Science **153:** 699.

Finnerty, J. R., and M. Q. Martindale, 1998 The evolution of the Hox cluster: insights from outgroups. Current Opinion in Genetics & Development **8:** 681-687.

Fisher, R. A., 1928 The possible modification of the response of the wild type to recurrent mutations. Am. Nat. **62:** 115-126.

Fisher, R. A., 1930 *The Genetical Theory of Natural Selection.* Clarendon Press, Oxford.

Fisher, R. A., 1958 Retrospect of criticisms of natural selection, pp. 84-98 in *Evolution as a Process*, edited by J. Huxley, A. C. Hardy and E. B. Ford. George Allen & Unwin Ltd, London.

Force, A., M. Lynch, F. B. Pickett, A. Amores, Y. L. Yan *et al.*, 1999 Preservation of duplicate genes by complementary, degenerative mutations. Genetics **151:** 1531-1545.

Frank, S. A., 1996 The design of natural and artificial adaptive systems, pp. 451-505 in *Adaptation*, edited by M. R. Rose, and G. V. Lauder. Academic Press, San Diego, California, USA.

Fu, Y. X., and W. H. Li, 1993 Statistical tests of neutrality of mutations. Genetics **133:** 693-709.

Futuyma, D. J., 1998 *Evolutionary Biology.* Sinauer, Sunderland, Massachusetts, USA.

Galtier, N., and J. R. Lobry, 1997 Relationships between genomic G+C content, RNA secondary structures, and optimal growth temperature in prokaryotes. J. Mol. Evol. **44:** 632-636.

Garrick, D., S. Fiering, D. I. K. Martin and E. Whitelaw, 1998 Repeat-induced gene silencing in mammals. Nature Genet. **18:** 56-59.

Gentles, A. J., and S. Karlin, 1999 Why are human G-protein-coupled receptors predominantly intronless? Trends Genet. **15:** 47-49.

Gibbs, P. E. M., W. F. Witke and A. Dugaiczyk, 1998 The molecular clock runs at different rates among closely related members of a gene family. J. Mol. Evol. **46:** 552-561.

Gibson, T. J., and J. Spring, 1998 Genetic redundancy in vertebrates: polyploidy and persistence of genes encoding multidomain proteins. Trends Genet. **14:** 46-49.

Gillespie, J. H., 1991 *The Causes of Molecular Evolution.* Oxford University Press, Oxford.

Gillespie, J. H., 1998 *Population genetics: a concise guide.* The John Hopkins University Press, London.

Golding, G. B., and A. M. Dean, 1998 The structural basis of molecular adaptation. Mol. Biol. Evol. **15:** 355-369.

Goldman, N., 1998 Effects of sequence alignment procedures on estimates of phylogeny. BioEssays **20:** 287-290.

Goldstein, D. B., 1992 Heterozygote advantage and the evolution of a dominant diploid phase. Genetics **132:** 1195-1198.

Goldstein, D. B., 1994 Deleterious mutations and the evolution of male haploidy. Am. Nat. **144:** 176-183.

Goldstein, D. B., and P. H. Harvey, 1999 Evolutionary inference from genomic data. Bioessays **21:** 148-156.

Gould, S. J., 1997 The exaptive excellence of spandrels as a term and prototype. Proc. Natl. Acad. Sci. USA **94:** 10750-10755.

Gould, S. J., and R. C. Lewontin, 1979 The spandrels of San Marco and the Panglossian paradigm: a critique of the adaptationist programme. Proc. R. Soc. Lond. B **205:** 581-598.

Green, R. F., and D. L. G. Noakes, 1995 Is a little bit of sex as good as a lot? J theor Biol **174:** 87-96.

Haig, D., 1993 The evolution of unusual chromosomal systems in coccoids: extraordinary sex ratios revisited. J. Evol. Biol. **6:** 69-77.

Haig, D., 1997 Parental antagonism, relatedness asymmetries, and genomic imprinting. Proc.R. Soc. Lond. B **264:** 1657-1662.

Haig, D., and C. Graham, 1991 Genomic imprinting and the strange case of the insulin-like growth factor II receptor. Cell **64:** 1045-6.

Haldane, J. B. S., 1927 A mathematical theory of natural and artificial selection. Part V. Selection and mutation. Proc. Camb. Phil. Soc. **23:** 838-844.

Haldane, J. B. S., 1930 A note on Fisher's theory of the origin of dominance and a correlation between dominance and linkage. Am. Nat. **64:** 87-90.

Haldane, J. B. S., 1932 *The Causes of Evolution.* Cornell University Press, New York.

Haldane, J. B. S., 1937 The effect of variation on fitness. Am. Nat. **71:** 337-349.

Haldane, J. B. S., 1957 The cost of natural selection. J. Genet. **55:** 511-524.

Haldane, J. B. S., 1964 A defense of beanbag genetics. Perspect. Biol. Med. **7:** 343-359.

Hall, B. G., 1998 Adaptive mutagenesis: a process that generates almost exclusively beneficial mutations. Genetica **103:** 109-125.

Hamilton, W., 1993 Inbreeding in Egypt and in this book : a childish view, pp. in *The Natural History of Inbreeding and Outbreeding*, edited by N. W. T. a. W. M. Shields. University of Chicago Press, Chicago.

Hamilton, W. D., 1967 Extraordinary sex ratios. Science **156:** 477-488.

Hamilton, W. D., 1979 Wingless and fighting males in fig wasps and other insects., pp. 167-220 in *Reproductive Competition, Mate Choice and Sexual Selection*, edited by M. S. Blum, and N. A. Blum. Academic Press.

Hamilton, W. D., 1995 *Narrow roads of gene land 1: evolution of social behaviour.* W.H.Freeman,

Hartl, D., 1972 A fundamental theorem of natural selection for sex linkage or arrhenotoky. Am. Zool. **11:** 309-325.

Hartl, D. L., and S. W. Brown, 1970 The origin of male haploid genetic systems and their expected sex ratio. Theor. Popul. Biol. **1:** 165-190.

Hartl, D. L., E. N. Moriyama and S. A. Sawyer, 1994 Selection intensity for codon bias. Genetics **138:** 227-234.

Harvey, P. H., and M. D. Pagel, 1991 *The Comparative Method in Evolutionary Biology.* Oxford University Press, Oxford.

Harvey, P. H., and L. Partridge, 1984 When deviants are favored - evolution of sex determination. Nature **307:** 689-691.

Hastings, I. M., 1992 Population genetic-aspects of deleterious cytoplasmic genomes and their effect on the evolution of sexual reproduction. Gen. Res. **59:** 215-225.

Hastings, I. M., 1994 Manifestations of sexual selection may depend on the genetic basis of sex determination. Proc. R. Soc. Lond. B **258:** 83-87.

Hayward, B. E., 1998 Bidirectional imprinting of a single gene: GNAS1 encodes maternally, paternally and biallelically derived proteins. Proc. Natl. Acad. Sci. USA `**9:** 15475-15480.

Hedrick, P. W., 1999 Antagonistic pleiotropy and genetic polymorphism: a perspective. Heredity **82:** 126-133.

Hedrick, P. W., and J. D. Parker, 1997 Evolutionary genetics and genetic variation of haplodiploids and X- linked genes. Annu. Rev. Ecol. Syst. **28:** 55-83.

Herbert, A., and A. Rich, 1999 RNA processing and the evolution of eukaryotes. Nature Genet. **21:** 265-269.

Herzel, H., E. N. Trifonov, O. Weiss and I. Grosse, 1998 Interpreting correlations in biosequences. Physica a **249:** 449-459.

Hey, J., 1999 The neutralist, the fly and the selectionist. TREE **14:** 35-38.

Hickey, D. A., 1993 Molecular Symbionts and the Evolution of Sex. J. Hered. **84:** 410-414.

Hill, W. G., and A. Robertson, 1966 The effect of linkage on limits to natural selection. Genet. Res. **8:** 269-294.

Hoekstra, R. F., 1987 The evolution of sexes, pp. 59-91 in *The evolution of sex and its consequences*, edited by S. C. Stearns. Birkhauser, Basil.

Holland, P. W. H., 1998 Major transitions in animal evolution: a developmental genetic perspective. Amer. Zool. **38:** 829-842.

Holland, P. W. H., and J. GarciaFernandez, 1996 Hox genes and chordate evolution. Developmental Biology **173:** 382-395.

Horowitz, N. H., 1945 On the evolution of biochemical syntheis. Proc, Natl. Acad. Sci. USA **31:** 153-157.

Hsu, T. S., 1975 A possible function of constitutive heterochromatin: the bodygaurd hypothesis. Genetics **79 S2:** 137-150.

Hudson, R. R., M. Kreitman and M. Agaude, 1987 A test of neutral molecular evolution based on nucleotide data. Genetics **116:** 153-159.

Hughes, A. L., 1994 The evolution of functionally novel proteins after gene duplication. Proceedings Of the Royal Society Of London Series B-Biological Sciences **256:** 119-124.

Hughes, A. L., 1998 Phylogenetic tests of the hypothesis of block duplication of homologous genes on human chromosomes 6, 9, and 1. Mol. Biol. Evol. **15:** 854-870.

Hughes, A. L., and M. Nei, 1988 Pattern of nucleotide substitution at major histocompatibility complex class I loci: evidence for overdominant selection. Nature **335:** 167-170.

Hughes, A. L., and M. Yeager, 1997 Comparative evolutionary rates of introns and exons in murine rodents. J. Mol. Evol. **45:** 125-130.

Hughes, M. K., and A. L. Hughes, 1993 Evolution of duplicate genes in a tetraploid animal, *Xenopus laevis*. Mol. Biol. Evol. **10:** 1360-1369.

Hurst, G. D. D., L. D. Hurst and R. Johnstone, 1992 Intra-nuclear conflict and its role in evolution. Trends Ecol. Evol. **7:** 373-378.

Hurst, G. D. D., L. D. Hurst and M. E. N. Majerus, 1997 Cytoplasmic sex ratio distorters, pp. 125-154 in *Influential Passengers*, edited by A. Hoffmann, S. O'Neill and J. Werren. Oxford University Press, Oxford.

Hurst, G. D. D., F. M. Jiggins, J. H. G. Von Schulenburg, D. Bertrand, S. A. West *et al.*, 1999 Male-killing *Wolbachia* in two species of insect. Proc. R. Soc. Lond. B. **266:** 735-740.

Hurst, L. D., 1990 Parasite diversity and the evolution of diploidy, multicellularity and anisogamy. J. theor. Biol. **144:** 429-443.

Hurst, L. D., 1991a The incidences and evolution of cytoplasmic male killers. Proc. R. Soc. Lond. B **244:** 91-99.

Hurst, L. D., 1991b Sex, slime and selfish genes. Nature **354:** 23-24.

Hurst, L. D., 1992a Intragenomic conflict as an evolutionary force. Proc. R. Soc. Lond. B **248:** 135-140.

Hurst, L. D., 1992b Is *Stellate* a relict meiotic driver? Genetics **130:** 229-230.

Hurst, L. D., 1993 The incidences, mechanisms and evolution of cytoplasmic sex ratio distorters in animals. Biol. Rev. **68:** 121-193.

Hurst, L. D., 1996 Further evidence consistent with *Stellate*'s involvement in meiotic drive. Genetics **142:** 641-643.

Hurst, L. D., 1997 Evolutionary theories of genomic imprinting, pp. 211-237 in *Frontiers in Molecular Biology: Genomic Imprinting in Mammals*, edited by W. Reik, and A. Surani. Oxford University Press, Oxford.

Hurst, L. D., 1998 Peromysci, promiscuity and imprinting. Nature Genet. **20:** 315-316.

Hurst, L. D., 1999 The evolution of genomic anatomy. Trends Ecol. Evol. **14:** 108-112.

Hurst, L. D., A. Atlan and B. O. Bengtsson, 1996a Genetic Conflicts. Q. Rev. Biol. **71:** 317-364.

Hurst, L. D., and H. Ellergren, 1998 Sex biases in the mutation rate. Trends Genet. **14:** 446-452.

Hurst, L. D., and W. D. Hamilton, 1992 Cytoplasmic fusion and the nature of sexes. Proc. R. Soc. Lond. B **247:** 189-194.

Hurst, L. D., G. McVean and T. Moore, 1996b Imprinted genes have few and small introns. Nature Genet. **12:** 234-237.

Hurst, L. D., and G. T. McVean, 1996 Clade selection, reversible evolution and the persistence of selfish elements - the evolutionary dynamics of cytoplasmic incompatibility. Proc. R. Soc. Lond. B **263:** 97-104.

Hurst, L. D., and G. T. McVean, 1997 Growth effects of uniparental disomies and the conflict theory of genomic imprinting. Trends Genet. **13:** 436-443.

Hurst, L. D., and G. T. McVean, 1998 Do we understand the evolution of genomic imprinting? Curr. Opin. Genet. Dev. **8:** 701-708.

Hurst, L. D., and P. Nurse, 1991 A note on the evolution of meiosis. J. theor. Biol. **150:** 561-563.

Huxley, C., and M. Fried, 1990 The mouse surfeit locus contains a cluster of six genes associated with four CpG rich islands in 32 kilobases of genomic DNA. Mol. Cell. Biol. **10:** 605-614.

Huynen, M. A., and P. Bork, 1998 Measuring genome evolution. Proc. Natl. Acad. Sci. USA **95:** 5849-5856.

Ikemura, T., 1985 Codon usage and tRNA content in unicellular and multicellular organisms. Mol. Biol. Evol. **2:** 13-34.

Iwasa, Y., 1998 The conflict theory of genomic imprinting: How much can be explained? Curr. Top. Dev. Biol. **40:** 255-293.

Jacob, F., and J. Monod, 1962 On the regulation of gene activity. Cold Spring Harbor Symp. Quant. Biol. **26:** 193-211.

Jenkins, C. D., 1993 Selection and the evolution of genetic life-cycles. Genetics **133:** 401-410.

Jenkins, D. L., C. A. Ortori and J. F. Y. Brookfield, 1995 A test for adaptive change in DNA sequences controlling transcription. Proc. R. Soc. Lond. B **261:** 203-207.

Jensen, S., M.-P. Gassama and T. Heidmann, 1999 Taming of transposable elements by homology depedent gene silencing. Nature Genet. **21:** 209-212.

Jockusch, E. L., 1997 An evolutionary correlate of genome size change in plethontid salamanders. Proc. R. Soc. Lond. B **264:** 597-604.

John, B., and G. L. G. Miklos, 1988 *The Eukaryote Genome in Development and Evolution.* Allen and Unwin, London.

Jones, P. A., 1999 The DNA methylation paradox. Trends Genet. **15:** 34-37.

Kacser, H., and J. A. Burns, 1981 The molecular basis of dominance. Genetics **97:** 639-666.

Karlin, S., and J. Mrazek, 1996 What drives codon choices in human genes? J. Mol. Biol. **262:** 459-472.

Kass, S. U., D. Pruss and A. P. Wolffe, 1997 How does DNA methylation repress transcription? Trends Genet. **13:** 444-449.

Katsanis, N., J. Fitzgibbon and E. M. C. Fisher, 1996 Paralogy mapping: Identification of a region in the human MHC triplicated onto human chromosomes 1 and 9 allows the prediction and isolation of novel PBX and NOTCH loci. Genomics **35:** 101-108.

Keightley, P. D., 1994 The distribution of mutation effects on viability in Drosophila melanogaster. Genetics **138:** 1315-1322.

Keller, N. P., and T. M. Hohn, 1997 Metabolic pathway gene clusters in filamentous fungi. Fungal Genet. Biol. **21:** 17-29.

Kendrew, J. (Editor), 1994 *The Encyclopedia of Molecular Biology.* Blackwell, Oxford.

Kimura, M., 1967 On the evolutionary adjustment of mutation rates. Genet. Res. **9:** 23-34.

Kimura, M., 1968a Evolutionary rate at the molecular level. Nature **217:** 624-626.

Kimura, M., 1968b Genetic variability maintained in a finite population due to mutational production of neutral and nearly neutral isoalleles. Genet. Res. **11:** 247-269.

Kimura, M., 1983 *The Neutral Theory of Evolution.* Cambridge University Press, Cambridge.

Kimura, M., and J. L. King, 1979 Fixation of a deleterious allele at one of two "duplicate" loci by mutation pressure and random drfit. Proc. Natl. Acad. Sci. USA **76:** 2858-2861.

Kimura, M., and T. Maruyama, 1966 The mutational load with epistatic gene interactions in fitness. Genetics **54:** 1337-1351.

Kimura, M., and T. Ohta, 1974 On some principles governing molecular evolution. Proc. Natl. Acad. Sci. USA **71:** 2848-2852.

King, J. L., and T. H. Jukes, 1969 Non-Darwinian evolution. Science **164:** 788-798.

Kirby, D. A., S. V. Muse and W. Stephan, 1995 Maintenance of pre-mRNA secondary structure by epistatic selection. Proc. Natl. Acad. Sci. USA **92:** 9047-9051.

Kirkpatrick, M., 1996 Genes and adaptation: a pocket guide to the theory, pp. 125-146 in *Adaptation*, edited by M. R. Rose, and G. V. Lauder. Academic Press, San Diego, California, USA.

Klaff, P., D. Riesner and G. Steger, 1996 RNA structure and the regulation of gene expression. Plant Mol. Biol. **32:** 89-106.

Kliman, R. M., and A. Eyre-Walker, 1998 Patterns of base composition within the genes of *Drosophila melanogaster*. J. Mol. Evol. **46:** 534-541.

Kliman, R. M., and J. Hey, 1993 Reduced natural selection associated with low recombination in *Drosophila melanogaster*. Mol. Biol. Evol. **10:** 1239-1258.

Klinger, T., 1993 The persistence of haplodiploidy in algae. Trends Ecol. Evol. **8:** 256-258.

Ko, M. S. H., T. A. Threat, X. Q. Wang, J. H. Horton, Y. S. Cui *et al.*, 1998 Genome-wide mapping of unselected transcripts from extraembryonic tissue of 7.5-day mouse embryos reveals enrichment in the t-complex and under-representation on the X chromosome. Hum. Mol. Genet. **7:** 1967-1978.

Kondrashov, A., 1982 Selection against harmful mutations in large sexual and asexual populations. Genet. Res. Camb. **40:** 325-332.

Kondrashov, A. S., 1988 Deleterious mutations and the evolution of sexual reproduction. Nature **336:** 435-440.

Kondrashov, A. S., 1994 The asexual ploidy cycle and the origin of sex. Nature **370:** 213-216.

Kondrashov, A. S., 1998 Measuring spontaneous deleterious mutation process. Genetica **103:** 183-197.

Kondrashov, A. S., and J. F. Crow, 1991 Haploidy or Diploidy - Which Is Better? Nature **351:** 314-315.

Korol, A. B., I. A. Preigel and S. I. Preigel, 1994 *Recombination Variability and Evolution.* Chapman and Hall, London.

Kreitman, M., 1996 The neutral theory is dead - long live the neutral theory. Bioessays **18:** 678-683.

Kreitman, M., and R. R. Hudson, 1991 Inferring the evolutionary histories of the Adh and Adh-dup loci in Drosophila melanogaster from patterns of polymorphism and divergence. Genetics **127:** 565-582.

Kricker, M. C., J. W. Drake and M. Radman, 1992 Duplication-targeted DNA methylation and mutagenesis in the evolution of eukaryotic chromosomes. Proc. Natl. Acad. Sci. USA **89:** 1075-1079.

Kundera, M., 1984 *The unbearable lightness of being.* Faber and Faber, London.

Lahn, B. T., and D. C. Page, 1997 Functional coherence of the human Y chromosome. Science **278:** 675-680.

Lauder, G. V., 1996 The argument from design, pp. 55-92 in *Adaptation*, edited by M. R. Rose, and G. V. Lauder. Academic Press, San Diego, California, USA.

Lawrence, J. G., and J. R. Roth, 1996 Selfish operons: horizontal transfer may drive the evolution of gene clusters. Genetics **143:** 1843-1860.

Lee, S. J., 1991 Expression of growth/differentiation factor 1 in the nervous system: conservation of a bicistronic structure. Proc. NAtl. Acad. Sci. USA **88:** 4250-4254.

Leicht, B. G., S. V. Muse, M. Hanczye and G. Clark, 1995 Constraints on intron evolution in the gene encoding the myosin alkali light chain in *Drosophila.* Genetics **139:** 299-308.

Leigh, E. G., 1970 Natural selection and mutability. Am. Nat. **104:** 301-305.

Leigh, E. G., 1973 The evolution of mutation rates. Genetics **73 S:** 1-18.

Leigh, E. G. J., 1991 Genes, bees and ecosystems: the evolution of a common interest among individuals. Trends Ecol. Evol. **6:** 257-262.

Lewin, B., 1997 *Genes VI.* Oxford University Press, Oxford.

Lewis, W. M., 1985 Nutrient scarcity as an evolutionary cause of haploidy. Am. Nat. **125:** 692-701.

Lewontin, R. C., 1993 *The Doctrine of DNA.* Penguin Books, London.

Li, W.-H., 1983 Evolution of duplicate genes and pseduogenes, pp. in *Evolution of Genes and Proteins*, edited by M. Nei, and R. K. Koehn. Sinauer Associates, Sunderland, MA.

Li, W. H., 1993 Unbiased estimation of the rates of synonymous and nonsynonymous substitution. J. Mol. Evol. **36:** 96-99.

Li, W. H., 1997 *Molecular Evolution.* Sinauer Associates, Sunderland, Massachusetts, USA.

Li, W. H., C. I. Wu and C. C. Luo, 1985a A new method for estimating synonymous and nonsynonymous rates of nucleotide substitution considering the relative likelihood of nucleotide and codon changes. Mol. Biol. Evol. **2:** 150-174.

Li, W. H., C.-I. Wu and C.-C. Luo, 1985b Evolution of DNA sequences, pp. 1-94 in *Molecular Evolutionary Genetics*, edited by R.J.MacIntyre. Plenum, New York.

Li, W.-H., and D. Graur, 1991 *Fundamentals of Molecular Evolution.* Sinauer associates, Sunderland, Massachusetts.

Lloyd, V. K., D. A. Sinclair and T. A. Grigliatti, 1999 Genomic imprinting and position-effect variegation in Drosophila melanogaster. Genetics **151:** 1503-1516.

Logsdon, J. M., 1998 The recent origins of spliceosomal introns revisited. Curr. Opin. Genet. Dev. **8:** 637-648.

Lundin, L. G., 1993 Evolution of the vertebrate genome as reflected in paralogous chromosomal regions in man and the house mouse. Genomics **16:** 1-9.

Lyttle, T. W., 1991 Segregation distorters. Annu. Rev. Genet. **25:** 511-557.

Mable, B. K., and S. P. Otto, 1998 The evolution of life cycles with haploid and diploid phases. Bioessays **20:** 453-462.

Majerus, T. M. O., M. E. N. Majerus, B. Knowles, J. Wheeler, D. Bertrand *et al.*, 1998 Extreme variation in the prevalence of inherited male-killing microorganisms between three populations of Harmonia axyridis (Coleoptera: Coccinellidae). Heredity **81:** 683-691.

Makalowski, W., J. Zhang and M. S. Boguski, 1996 Comparative analysis of 1196 orthologous mouse and human full-length mRNA and protein sequences. Genome Res. **6:** 846-857.

Maniatis, T., E. F. Fritsch, J. Lauer and R. M. Lawn, 1980 The molecular genetics of human hemoglobins. Annu. Rev. Genet. **14:** 145-178.

Masterson, J., 1994 Stomatal size in fossil plants - evidence for polyploidy in majority of angiosperms. Science **264:** 421-424.

Maynard Smith, J., 1978 *The Evolution of Sex.* Cambridge University Press, Cambridge.

Maynard Smith, J., and J. Haigh, 1974 The hitch-hiking effect of a favourable gene. Genet. Res. **23:** 23-35.

Maynard Smith, J., and N. H. Smith, 1996 Site-specific codon bias in bacteria. Genetics **142:** 1037-1043.

Maynard Smith, J., and E. Szathmary, 1995 *The Major Transitions in Evolution.* W.H. Freeman, Spektrum, Oxford.

Mayr, E., 1955 Integration of genotypes: synthesis. Cold Harbor Symp. Quant. Biol. **20:** 327-333.

Mayr, E., 1963 *Animal species and evolution.* Harvard University Press, Cambridge, USA.

Mazat, J. P., C. Reder and T. Letellier, 1996 Why are most flux control coefficients so small? Journal of Theoretical Biology **182:** 253-258.

McAdams, H. H., and A. Arkin, 1999 It's a noisy business! Genetic regulation at the nanomolar scale. Trends Genet. **15:** 65-69.

McDonald, J., and M. Kreitman, 1991 Adaptive protein evolution at *Adh* locus in *Drosophila.* Nature **351:** 652-654.

McVean, G. T., 1997 Adaptation and conflict: the differences between the sexes in mammalian genome evolution, PhD Thesis, Department of Genetics, University of Cambridge, Cambridge.

McVean, G. T., and L. D. Hurst, 1997a Evidence for a selectively favourable reduction in the mutation rate of the X chromosome. Nature **386:** 388-392.

McVean, G. T., and L. D. Hurst, 1997b Molecular evolution of imprinted genes: No evidence for antagonistic coevolution. Proc. R. Soc. Lond. B **264:** 739-746.

Miyata, T., H. Hayashida, K. Kuma, K. Mitsuyasu and T. Yasunaga, 1987 Male-driven molecular evolution: a model and nucleotide sequence analysis. Cold Spring Harbor Symp. Quant. Biol. **52:** 863-867.

Mochizuki, A., Y. Takeda and Y. Iwasa, 1996 The evolution of genomic imprinting. Genetics **144:** 1283-1295.

Moore, T., and D. Haig, 1991 Genomic imprinting in mammalian development: a parental tug-of-war. Trends Genet. **7:** 45-9.

Morgan, D. O., J. C. Edman, D. N. Standring, V. A. Fried, M. C. Smith *et al.*, 1987 Insulin-like growth factor-II receptor as a multifunctional binding- protein. Nature **329:** 301-307.

Moriyama, E. N., and J. R. Powell, 1997 Codon usage bias and tRNA abundance in Drosophila. J. Mol. Evol. **45:** 514-523.

Moriyama, E. N., and J. R. Powell, 1998 Gene length and codon usage bias in *Drosophila melanogaster*, *Saccharomyces cerevisiae* and *Escherichia coli*. Nucleic Acid Res. **26:** 3188-3193.

Mouchiroud, D., C. Gautier and G. Bernardi, 1995 Frequencies of synonymous substitutions in mammals are gene-specific and correlated with frequencies of nonsynonymous substitutions. J. Mol. Evol. **40:** 107-113.

Mount, S. M., C. Burks, G. Hertz, G. D. Stormo, O. White *et al.*, 1992 Splicing signals in Drosophila: intron size, information content, and consensus sequences. Nucleic Acids Res. **20:** 4255-4262.

Muller, H. J., 1925 Why polyploidy is rarer in animals than in plants. Am. Nat. **59:** 346-353.

Muller, H. J., 1950 Our load of mutations. Am. J. Hum. Genet. **2:** 111-176.

Nachman, M. W., and G. A. Churchill, 1996 Heterogeneity in rates of recombination across the mouse genome. Genetics **142:** 537-548.

Nadeau, J. H., and D. Sankoff, 1997 Comparable rates of gene loss and functional divergence after genome duplications early in vertebrate evolution. Genetics **147:** 1259-1266.

Nagy, E., and L. E. Maquat, 1998 A rule for termination-codon position within intron-containing genes: when nonsense affects RNA abundance. Trends Biochem. Sci. **23:** 198-199.

Nei, M., 1969 Gene duplication and nucleotide substitution in evolution. Nature **221:** 40-42.

Neumann, B., P. Kubicka and D. P. Barlow, 1995 Characteristics of imprinted genes. Nature Genet. **9:** 12-13.

Neumann, B., A. Wutz, O. W. Smrzka, D. P. Barlow and R. Lyle, 1997 Imprinting at the mouse and human IGF2R loci, pp. 38-52 in *Genomic imprinting*, edited by W. Reik, and A. Surani. Oxford University Press, Oxford.

Nothel, H., 1987 Adaptation of *Drosophila melanogaster* populations to high mutation pressure: evolutionary adjustment of mutation rates. Proc. Nat. Acad. Sci. USA **84:** 1045-1049.

Nur, U., 1980 Evolution of unusual chromosome systems in scale insects (Coccoidea: Homoptera), pp. in *Insect Cytogentics*, edited by R. L. Blackman, G. M. Hewitt and M. Ashburner. Blackwell Scientific Publications, Oxford.

Ohlsson, R., B. Tycko and C. Sapienza, 1998 Monoallelic expression: 'there can only be one'. Trends Genet. **14:** 435-438.

Ohno, S., 1970 *Evolution by gene duplication.* Springer-Verlag, Heidelberg.

Ohno, S., 1972 So much "junk" DNA in our genome, in *Evolution of Genetic Systems*. Brookhaven Symp. Biol.

Ohno, S., 1973 Ancient linkage groups and frozen accidents. Nature **244:** 259-262.

Ohta, T., 1972 Population size and rate of evolution. J. Mol. Evol. **1:** 305-314.

Ohta, T., 1989 The mutational load of a multigene family with uniform members. Genet. Res. **53:** 141-145.

Ohta, T., 1993 An examination of the generation time effect on molecular evolution. Proc. Natl. Acad. Sci. USA **90:** 10676-10680.

Ohta, T., 1995 Synonymous and nonsynonymous substitutions in mammalian genes and the nearly neutral theory. J. Mol. Evol. **40:** 56-63.

Ohta, T., and Y. Ina, 1995 Variation in synonymous substitution rates among mammalian genes and the correlation between synonymous and nonsynonymous divergences. J. Mol. Evol. **41:** 717-720.

Oliver, J. H., 1971 Parthenogenesis in mites and ticks (Arachnida: Acari). Amer. Zool. **11:** 283-299.

Orgel, L. E., and F. H. C. Crick, 1980 Selfish DNA: the ultimate parasite. Nature **284:** 604-607.

Orr, H. A., 1990 "Why polyploidy in rarer in animals than in plants revisited" revisited. Am. Nat. **136:** 759-770.

Orr, H. A., 1991 A test of Fisher's theory of dominance. Proc. Natl. Acad. Sci. USA **88:** 11413-11415.

Orr, H. A., 1995 Somatic mutation favors the evolution of diploidy. Genetics **139:** 1441-1447.

Orr, H. A., and S. P. Otto, 1994 Does diploidy increase the rate of adaptation? Genetics **136:** 1475-1480.

Ostergren, G., 1945 Parasitic nature of extra fragment chromosomes. Bot. Notiser **2:** 157-163.

Otto, S. P., 1997 Unravelling gene interactions. Nature **390:** 343-344.

Otto, S. P., and D. B. Goldstein, 1992 Recombination and the evolution of diploidy. Genetics **131:** 745-751.

Otto, S. P., and J. C. Marks, 1996 Mating systems and the evolutionary transition between haploidy and diploidy. Biol. J. Linn. Soc. **57:** 197-218.

Otto, S. P., and M. E. Orive, 1995 Evolutionary consequences of mutation and selection within an individual. Genetics **141:** 1173-1187.

Pagel, M., 1999 Mother and father in surprise genetic agreement. Nature **397:** 19-20.

Pagel, M. D., and R. A. Johnstone, 1992 Variation across species in the size of the nuclear genome supports the junk-DNA explanation for the C-value paradox. Proc. R. Soc. Lond. B **249:** 119-124.

Pallavicini, A., R. Zimbello, N. Tiso, T. Muraro, L. Rampoldi *et al.*, 1997 The preliminary transcript map of a human skeletal muscle. Hum. Mol. Genet. **6:** 1445-1450.

Parsons, P. A., and W. F. Bodmer, 1961 The evolution of overdominance: natural selection and heterozygote advantage. Nature **190:** 7-12.

Partridge, L., and L. D. Hurst, 1998 Sex and conflict. Science **281:** 2003-2008.

Perrot, V., S. Richerd and M. Valero, 1991 Transition From Haploidy to Diploidy. Nature **351:** 315-317.

Pesole, G., G. Dellisanti, G. Preparata and C. Saccone, 1995 The importance of base composition in the correct assessment of genetic distance. J. Mol. Evol. **41:** 1124-1127.

Petrov, D. A., and D. L. Hartl, 1998 High rate of DNA loss in the Drosophila melanogaster and Drosophila virilis species groups. Mol. Biol. Evol. **15:** 293-302.

Petrov, D. A., E. R. Lozovskaya and D. L. Hartl, 1996 High intrinsic rate of DNA loss in Drosophila. Nature **384:** 346-349.

Philippe, H., and J. Laurent, 1998 How good are deep phylogenetic trees? Curr. Opin. Genet. Dev. **8:** 616-623.

Piatigorsky, J., and G. J. Wistow, 1989 Enzyme / crystallins: gene sharing as an evolutionary strategy. Cell **57:** 197-199.

Porteous, J. W., 1996 Dominance - 115 years after Mendel's paper. J. theor. Biol. **182:** 223-232.

Powell, J. R., 1997 *Progress and prospects in evolutionary biology: the Drosophila model.* Oxford University Press, New York.

Powell, J. R., and E. N. Moriyama, 1997 Evolution of codon usage bias in Drosophila. Proc. Natl. Acad. Sci. USA **94:** 7784-7790.

Price, G. R., 1970 Selection and covariance. Nature **227:** 520-521.

Ptacek, M. B., H. C. Gerhardt and R. D. Sage, 1994 Speciation by polyploidy in treefrogs: multiple origins of the tetraploid, *Hyla versicolor*. Evolution **48:** 898-908.

Purvis, I. J., A. J. E. Bettany, T. C. Santiago, J. R. Coggins, K. Duncan *et al.*, 1987 The efficiency of folding of some proteins is increased by controlled rates of translation *in vivo*: a hypothesis. J. Mol. Biol. **193:** 413-417.

Raikov, I. B., 1982 *The Protozoan Nucleus.* Springer, Berlin.

Reeve, H. K., and P. W. Sherman, 1993 Adaptation and the goals of evolutionary research. Q. Rev. Biol. **68:** 1-32.

Reik, W., and E. A. Maher, 1997 Imprinting in clusters: lessons from Beckwith-Wiedemann syndrome. Trends Genet. **13:** 330-334.

Reinhold, K., 1998 Sex linkage among genes controlling sexually selected traits. Behav. Ecol. Socio. **44:** 1-7.

Rice, W. R., 1984 Sex chromosomes and the evolution of sexual dimorphism. Evolution **38:** 735-742.

Rice, W. R., 1987a The accumulation of sexually antagonistic genes as a selective agent promoting the evolution of reduced recombination between primitive sex chromosomes. Evolution **41:** 911-914.

Rice, W. R., 1987b Genetic hitchhiking and the evolution of reduced genetic activity of the Y sex chromosome. Genetics **116:** 161-7.

Rice, W. R., 1992 Sexually antagonistic genes - experimental evidence. Science **256:** 1436-1439.

Ridley, M., 1996 *Evolution.* Blackwell Science, Cambridge, Massachusetts, USA.

Rose, M. R., and G. V. Lauder (Editors), 1996 *Adaptation.* Academic Press, San Diego, California, USA.

Ruddle, F. H., 1997 Vertebrate genome evolution - The decade ahead. Genomics **46:** 171-173.

Ruddle, F. H., K. L. Bentley, M. T. Murtha and N. Risch, 1994 Gene loss and gain in the evolution of the vertebrates. Development 155-161.

Saifi, G. M., and H. S. Chandra, 1999 An apparent excess of sex- and reproduction-related genes on the human X chromosome. Proc. R. Soc. Lond. B **266:** 203-209.

Schafer, A., 1994 Genes and phenotypes of the human Y chromosome. Reprod. Med. Rev. **3:** 77-95.

Schwartz, R., and J. F. Curran, 1997 Analyses of frameshifting at UUU-pyrimidine sites. Nucleic Acids Res. **25:** 2005-2011.

Sharman, A. C., and P. W. H. Holland, 1996 Conservation, duplication, and divergence of developmental genes during chordate evolution. Netherlands J. Zool. **46:** 47-67.

Sharp, P. M., and W. H. Li, 1987 The rate of synonymous substitution in enterobacterial genes is inversely related to codon usage bias. Mol. Biol. Evol. **4:** 222-230.

Sharp, P. M., and W. H. Li, 1989 On the rate of DNA sequence evolution in *Drosophila.* J. Mol. Evol. **28:** 398-402.

Sharp, P. M., M. Stenico, J. F. Peden and A. T. Lloyd, 1993 Codon usage - mutational bias, translational selection, or both. Biochem. Soc. Trans. **21:** 835-841.

Sharp, P. M., T. M. F. Tuohy and K. R. Mosurski, 1986 Codon usage in yeast: cluster analysis clearly differentiates highly and lowly expressed genes. Nucleic Acid Res. **14:** 5125-5143.

Shibata, H., Y. Yoda, R. Kato, T. Ueda, M. Kamiya *et al.*, 1998 A methylation imprint mark in the mouse imprinted gene Grf1/Cdc25Mm locus shares a common feature with the U2afbp-rs gene: An association with a short tandem repeat and a hypermethylated region. Genomics **49:** 30-37.

Shields, D. C., P. M. Sharp, D. G. Higgins and F. Wright, 1988 Silent sites in Drosophila genes are not neutral - evidence of selection among synonymous codons. Mol. Biol. Evol. **5:** 704-716.

Sidow, A., 1996 Gen(om)e duplications in the evolution of early vertebrates. Curr. Opin. Genet. Dev. **6:** 715-722.

Simmons, M. J., and J. F. Crow, 1977 Mutations affecting fitness in *Drosophila* populations. Annu. Rev. Genet. **11:** 49-78.

Sinervo, B., and A. L. Basolo, 1996 Testing adaptation using phenotypic manipulations, pp. 149-185 in *Adaptation*, edited by M. R. Rose, and G. V. Lauder. Acadmeic Press, San Diego, California, USA.

Skrabanek, L., and K. H. Wolfe, 1998 Eukaryote genome duplication - where's the evidence. Curr. Opin. Genet. Dev. **8:** 694-700.

Smith, C. W. J., J. G. Patton and B. Nadal-Ginard, 1989 Alternative splicing in the control of gene expression. Annu. Rev. Genet. **23:** 527-577.

Sniegowski, P., 1997 Evolution: setting the mutation rate. Curr. Biol. **7:** R487-R488.

Sniegowski, P. D., P. J. Gerrish and R. E. Lenski, 1997 Evolution of high mutation rates in experimental populations of Escherichia coli. Nature **387:** 703-705.

Sparrow, A. H., and A. F. Nauman, 1976 Evolution of genome size by DNA doublings. Science **192:** 524-529.

Spencer, H. G., A. G. Clark and M. W. Feldman, 1999 Genetic conflicts and the evolutionary origin of genomic imprinting. Trends Ecol. Evol. **14:** 197-201.

Spencer, H. G., and M. J. M. Williams, 1997 The evolution of genomic imprinting: two modifier-locus models. Theor. Popul. Biol. **51:** 23-35.

Spring, J., 1997 Hypothesis vertebrate evolution by interspecific hybridisation - Are we polyploid? FEBS Letters **400:** 2-8.

Springer, M. S., L. J. Hollar and A. Burk, 1995 Compensatory substitutions and the evolution of the mitochondrial 12s rRNA gene in mammals. Mol. Biol. Evol. **12:** 1138-1150.

Sturtevant, A. H., 1937 Essays on evolution I. On the effects of selection on mutation rate. Q. Rev. Biol. **12:** 464-467.

Szathmáry, E., 1993 Do deleterious mutations act synergistically? Metaboic control theory provides a partial answer. Genetics **133:** 127-132.

Szathmary, E., I. Scheuring, M. Kotsis and I. Galdkih, 1990 Sexuality of eukaryotic unicells: hyperbolic growth, coexistence of facultative pathenogens, and the repair hypothesis, pp. 279-287 in *Organizational constraints on the dynamics of evolution*, edited by J. Maynard Smith, and G. Vida. Manchester University Press, Manchester.

Tajima, F., 1989 Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics **123:** 585-595.

Tanimoto, K., Q. Liu, J. Bungert and J. D. Engel, 1999 Effects of altered gene order or orientation of the locus control region on human B-globin gene expression in mice. Nature **398:** 344-348.

Thornhill, A. R., and P. S. Burgoyne, 1993 A paternally imprinted X chromosome retards the development of the early mouse embryo. Development **118:** 171-4.

Tinbergen, N., 1963 On aims and methods of ethology. Z. Tierpsychol. **20:** 410-433.

Trivers, R. L., 1974 Parent-offspring conflict. Am. Zool. **14:** 249-264.

Tsaur, S. C., and C. I. Wu, 1997 Positive selection and the molecular evolution of a gene of male reproduction, Acp26Aa of Drosophila. Mol. Biol. Evol. **14:** 544-549.

Valero, M., S. Richerd, V. Perrot and C. Destombe, 1992 Evolution of Alternation of Haploid and Diploid Phases in Life-Cycles. Trends Ecol. Evol. **7:** 25-29.

Vinogradov, A. E., 1998 Buffering: a possible passive-homeostasis role for redundant DNA. J. theor. Biol. **193:** 197-199.

Vrana, P. B., X. J. Guan, R. S. Ingram and S. M. Tilghman, 1998 Genomic imprinting is disrupted in interspecific Peromyscus hybrids. Nature Genet. **20:** 362-365.

Wagner, A., 1998a The fate of duplicated genes: loss or new function? Bioessays **20:** 785-788.

Wagner, A., 1999 Redundant gene functions and natural selection. J. Evol. Biol. **12:** 1-16.

Wagner, G., 1998b Complexity matters. Science **279:** 1158-1159.

Wallis, M., 1981 The molecular evolution of pituitary growth hormone prolactin and placental lactogen: a protein family showing variable rates of evolution. J. Mol. Evol. **17:** 10-18.

Wallis, M., 1993 Remarkably high rate of molecular evolution of ruminant placental lactogens. J. Mol. Evol. **37:** 86-88.

Waxman, D., and J. R. Peck, 1998 Pleiotropy and the preservation of perfection. Science **279:** 1210-1213.

Werren, J. H., 1993 The evolution of inbreeding in haplodiploid organisms, pp. 42-59 in *The natural history of inbreeding and outbreeding*, edited by N. W. Thornhill. University of Chicago Press, Chicago.

Werren, J. H., G. D. D. Hurst, W. Zhang, J. A. J. Breeuwer, R. Stouthamer *et al.*, 1994 Rickettsial relative associated with male killing in the ladybird beetle (*Adalia bipunctata*). J. Bact. **176:** 388-394.

Werren, J. H., U. Nur and C.-I. Wu, 1988 Selfish genetic elements. Trends Ecol. Evol. **3:** 297-302.

Werren, J. H., W. Zhang and L. R. Guo, 1995 Evolution and phylogeny of *Wolbachia* - reproductive parasites of arthropods. Proc. R. Soc. Lond. B **261:** 55-63.

Whiting, P. W., 1945 The evolution of male haploidy. Q. Rev. Biol. **20:** 231-260.

Widom, J., 1996 Short-range order in two eukaryotic genomes: relation to chromosome structure. J. Mol. Biol. **259:** 579-588.

Williams, G. C., 1992 *Natural Selection: Domains, Levels and Challenges.* Oxford University Press, New York.

Williamson, D. L., and D. F. Poulson, 1979 Sex ratio organisms (Spiroplasmas) of *Drosophila*, pp. 175-208 in *The Mycoplasmas*, edited by R. F. Whitcomb, and J. G. Tully. Academic Press, New York.

Wilson, A. C., S. S. Carlson and T. J. White, 1977 Biochemical evolution. Ann. Rev. Biochem. **46:** 573-639.

Wolfe, K. H., and P. M. Sharp, 1993 Mammalian gene evolution - nucleotide sequence divergence between mouse and rat. J. Mol. Evol. **37:** 441-456.

Wolfe, K. H., P. M. Sharp and W. H. Li, 1989 Mutation rates differ among regions of the mammalian genome. Nature **337:** 283-285.

Wolfe, K. H., and D. C. Shields, 1997 Molecular evidence for an ancient duplication of the entire yeast genome. Nature **387:** 708-713.

Wolffe, A. P., 1998 When more is less. Nature Genet. **18:** 5-6.

Wright, F., 1990 The effective number of codons used in a gene. Gene **87:** 23-29.

Wright, S., 1931 Evolution in Mendelian populations. Genetics **16:** 97-159.

Wright, S., 1932 The roles of mutation, inbreeding, cross-breeding and selection in evolution. Proc. VI Intern. Congr. Genet. **1:** 356-366.

Wright, S., 1934 Molecular and evolutionary theories of dominance. Am. Nat. **68:** 24-53.

Wu, C.-I., and M. F. Hammer, 1991 Molecular evolution of ultraselfish genes of meiotic drive systems, pp. 177-203 in *Evolution at the Molecular Level*, edited by R. K. Selander, A. G. Clark and T. S. Whittam. Sinauer Associates, Sunderland, Mass.

Wutz, A., O. W. Smrzka, N. Schweifer, K. Schellander, E. F. Wagner *et al.*, 1997 Imprinted expression of the Igf2r gene depends on an intronic CpG island. Nature **389:** 745-749.

Yang, Z., 1996 The among-site rate variation and its impact on phylogenetic analyses. Trends Ecol. Evol. **11:** 367-372.

Zakian, V. A., 1989 Structure and function of telomeres. Annu. Rev. Genet. **23:** 579-604.

Zharkikh, A., 1994 Estimation of evolutionary distances between nucleotide sequences. J. Mol. Evol. **39:** 315-329.

Zimmer, E. A., S. L. Martin, S. M. Beverley, Y. W. Kan and A. C. Wilson, 1980 Rapid duplication and loss of genes coding for the a chains of hemoglobin. Proc. Natl. Acad. Sci. USA **77:** 2158-2162.

Zoubak, S., G. Donofrio, S. Caccio, G. Bernardi and G. Bernardi, 1995 Specific Compositional Patterns of Synonymous Positions in Homologous Mammalian Genes. J. Mol. Evol. **40:** 293-307.

Zuckerkandl, E., 1976 Gene control in eukaryotes and the C value paradox: "excess" DNA as an impediment to transcription of coding sequences. J. Mol. Evol. **9:** 73-104.

Zuckerkandl, E., and W. Hennig, 1995 Tracking heterochromatin. Chromosoma **104:** 75-83.

Zuckerkandl, E., and L. Pauling, 1965 Evolutionary divergence and convergence in proteins, pp. 97-166 in *Evolving Genes and Proteins*, edited by V. Bryson, and H. J. Vogel. Academic Press, New York.